

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:46:51 ; Search time 9.85714 Seconds  
(without alignments)  
26.864 Million cell updates/sec

Title: US-09-700-329-1

Perfect score: 53

Sequence: 1 ECPWLEEE 9

Scoring table: BLOSUM62

Gapop 10.0 ,|gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2.6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2.6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2.6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2.6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2.6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2.6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	9	1	US-08-151-219-2
2	53	100.0	9	5	PCT-US94-13205-2
3	53	100.0	12	2	US-08-652-971-10
4	53	100.0	12	2	US-08-991-258A-10
5	53	100.0	12	2	US-08-769-399-10
6	53	100.0	12	3	US-08-991-953A-10
7	53	100.0	15	1	US-08-151-219-4
8	53	100.0	15	5	PCT-US94-13205-4
9	53	100.0	16	1	US-08-151-219-1
10	53	100.0	16	5	PCT-US94-13205-1
11	53	100.0	17	4	US-09-174-216-4
12	50	94.3	12	1	US-08-446-692-75
13	50	94.3	12	2	US-08-488-351A-75
14	50	94.3	17	1	US-08-446-692-74
15	50	94.3	17	2	US-08-488-351A-74
16	50	94.3	17	4	US-09-079-372-17
17	50	94.3	18	4	US-09-079-372-10
18	50	94.3	18	4	US-09-079-372-15
19	50	94.3	21	4	US-09-079-372-12
20	50	94.3	34	1	US-08-446-692-69
21	50	94.3	34	2	US-08-488-351A-69
22	50	94.3	34	4	US-09-079-372-9
23	50	94.3	34	4	US-09-079-372-16
24	50	94.3	35	4	US-09-079-372-8
25	50	94.3	47	1	US-08-446-692-100
26	50	94.3	47	2	US-08-488-351A-100
27	50	94.3	52	1	US-08-446-692-99

28 50 94.3 52 2 US-08-488-351A-99 Sequence 99, Appl  
29 50 94.3 69 1 US-08-446-692-95 Sequence 95, Appl  
30 50 94.3 69 2 US-08-488-351A-95 Sequence 95, Appl  
31 50 94.3 74 4 US-09-079-372-7 Sequence 7, Appl  
32 50 94.3 100 4 US-09-079-372-14 Sequence 14, Appl  
33 48 90.6 17 4 US-09-174-263-2 Sequence 2, Appl  
34 45 84.9 8 1 US-08-446-692-76 Sequence 76, Appl  
35 45 84.9 8 2 US-08-488-351A-76 Sequence 76, Appl  
36 45 84.9 17 1 US-08-127-351-3 Sequence 3, Appl  
37 45 84.9 17 1 US-08-480-367B-3 Sequence 3, Appl  
38 45 84.9 17 1 US-08-487-221A-3 Sequence 3, Appl  
39 45 84.9 17 1 US-08-480-370-3 Sequence 3, Appl  
40 39 73.6 470 1 US-08-416-478A-7 Sequence 7, Appl  
41 39 73.6 470 2 US-08-474-988B-7 Sequence 7, Appl  
42 39 73.6 470 2 US-08-394-442B-7 Sequence 7, Appl  
43 39 73.6 471 1 US-08-416-478A-2 Sequence 2, Appl  
44 39 73.6 471 2 US-08-474-988B-2 Sequence 2, Appl  
45 39 73.6 471 2 US-08-394-442B-2 Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-151-219-2  
; Sequence 2, Application US/08151219  
; Patent No. 5468494  
; GENERAL INFORMATION:  
; APPLICANT: Gevas, Phillip C.  
; APPLICANT: Grimes, Stephen  
; APPLICANT: Kart, Stephen  
; APPLICANT: Michaeli, Dov  
; APPLICANT: Scibienski, Robert  
; TITLE OF INVENTION: IMPROVED IMMUNOGENIC COMPOSITIONS  
; TITLE OF INVENTION: AGAINST HUMAN GASTRIN 17  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dimitrios T. Drivas, White and Case  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 100036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/151.219  
; FILING DATE: 12-NOV-1993  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Drivas, Dimitrios T.  
; REGISTRATION NUMBER: 32,218  
; REFERENCE/DOCKET NUMBER: 1102865-028  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 819-8286  
; TELEFAX: (212) 354-8113  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
US-08-151-219-2

Query Match 100.0%; Score 53; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9

RESULT 2  
PCT-US94-13205-2  
; Sequence 2, Application PC/TUS9413205  
; GENERAL INFORMATION:  
; APPLICANT: Gevas, Philip C.  
; APPLICANT: Grimes, Stephen  
; APPLICANT: Karr, Stephen  
; APPLICANT: Michaeli, Dov  
; APPLICANT: Scibielski, Robert  
; TITLE OF INVENTION: IMPROVED IMMUNOGENIC COMPOSITIONS  
; TITLE OF INVENTION: AGAINST HUMAN GASTRIN 17  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dimitrios T. Drivas, White and Case  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 100036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/13205  
; FILING DATE: 12-NOV-1993  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Drivas, Dimitrios T.  
; REGISTRATION NUMBER: 32,218  
; REFERENCE/DOCKET NUMBER: 1102865-028  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 819-8286  
; TELEFAX: (212) 814-8113  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
PCT-US94-13205-2

Query Match 100.0%; Score 53; DB 5; Length 9;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9

RESULT 3  
US-08-652-971-10  
; Sequence 10, Application US/08652971  
; Patent No. 5814507  
; GENERAL INFORMATION:  
; APPLICANT: Cheng, Jill  
; APPLICANT: Lasky, Laurence A.  
; TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.

STREET: 460 Point San Bruno Blvd.  
CITY: South San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,971  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: P1033  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 225-3216  
TELEFAX: (415) 952-9881  
TELEX: 910 371-7168  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-652-971-10

Query Match 100.0%; Score 53; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.0043;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9

RESULT 4  
US-08-991-258A-10  
; Sequence 10, Application US/08991258A  
; Patent No. 5928887  
; GENERAL INFORMATION:  
; APPLICANT: Cheng, Jill  
; APPLICANT: Lasky, Laurence A.  
; TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT, LLP  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/991,258A  
FILING DATE: 17-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/652,971  
FILING DATE: 24-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Walter H.  
REGISTRATION NUMBER: 24,190  
REFERENCE/DOCKET NUMBER: A-63478-3/WH/MTK

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX:  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-991-258A-10

Query Match 100.0%; Score 53; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.0043;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9

RESULT 5  
US-08-769-399-10

Sequence 10, Application US/08769399  
Patent No. 5976852  
GENERAL INFORMATION:  
APPLICANT: Cheng, Jill  
APPLICANT: Lasky, Laurence A.  
TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd.  
CITY: South San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,399  
FILING DATE:

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: P1033  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 225-3216  
TELEFAX: (415) 952-9881  
TELEX: 910 371-7168  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-769-399-10

Query Match 100.0%; Score 53; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.0043;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9

RESULT 6

US-08-991-953A-10  
Sequence 10, Application US/08991953A  
Patent No. 6083748  
GENERAL INFORMATION:  
APPLICANT: Cheng, Jill  
APPLICANT: Lasky, Laurence A.  
TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT, LLP  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/991,953A  
FILING DATE: 16-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/652,971  
FILING DATE: 24-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Walter H.  
REGISTRATION NUMBER: 24,190  
REFERENCE/DOCKET NUMBER: A-63478-3/WH/MTK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX:  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-991-953A-10

Query Match 100.0%; Score 53; DB 3; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.0043;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9

RESULT 7

US-08-151-219-4  
Sequence 4, Application US/08151219  
Patent No. 5468494  
GENERAL INFORMATION:  
APPLICANT: Gevas, Philip C.  
APPLICANT: Grimes, Stephen  
APPLICANT: Karr, Stephen  
APPLICANT: Michaeli, Dov  
APPLICANT: Scibienski, Robert  
TITLE OF INVENTION: IMPROVED IMMUNOGENIC COMPOSITIONS  
TITLE OF INVENTION: AGAINST HUMAN GASTRIN 17  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dimitrios T. Drivas, White and Case  
STREET: 1155 Avenue of the Americas  
CITY: New York

STATE: New York  
COUNTRY: U.S.A.  
ZIP: 100036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/151,219  
FILING DATE: 12-NOV-1993  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Drivas, Dimitrios T.  
REGISTRATION NUMBER: 32,218  
REFERENCE/DOCKET NUMBER: 1102865-028  
TELEPHONE: (212) 819-8286  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
US-08-151-219-4

Query Match 100.0%; Score 53; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.0055;  
Matches 9; Conservative 0; Mismatches 0; Indels 0;

OY 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9

RESULT 8  
PCT-US94-13205-4  
Sequence 4, Application PC/TUS9413205  
GENERAL INFORMATION:  
APPLICANT: Gevas, Philip C.  
APPLICANT: Grimes, Stephen  
APPLICANT: Kart, Stephen  
APPLICANT: Michaeli, Dov  
APPLICANT: Scibienski, Robert  
TITLE OF INVENTION: IMPROVED IMMUNOGENIC COMPOSITIONS  
TITLE OF INVENTION: AGAINST HUMAN GASTRIN 17  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dimitrios T. Drivas, White and Case  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 100036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/13205  
FILING DATE: 12-NOV-1993  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Drivas, Dimitrios T.  
REGISTRATION NUMBER: 32,218  
REFERENCE/DOCKET NUMBER: 1102865-028  
TELEPHONE: (212) 819-8286

TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
PCT-US94-13205-4

Query Match 100.0%; Score 53; DB 5; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.0055;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9

RESULT 9  
US-08-151-219-1  
Sequence 1, Application US/08151219  
Patent No. 5468494  
GENERAL INFORMATION:  
APPLICANT: Gevas, Philip C.  
APPLICANT: Grimes, Stephen  
APPLICANT: Kart, Stephen  
APPLICANT: Michaeli, Dov  
APPLICANT: Scibienski, Robert  
TITLE OF INVENTION: IMPROVED IMMUNOGENIC COMPOSITIONS  
TITLE OF INVENTION: AGAINST HUMAN GASTRIN 17  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dimitrios T. Drivas, White and Case  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 100036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/151,219  
FILING DATE: 12-NOV-1993  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Drivas, Dimitrios T.  
REGISTRATION NUMBER: 32,218  
REFERENCE/DOCKET NUMBER: 1102865-028  
TELEPHONE: (212) 819-8286  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
US-08-151-219-1

Query Match 100.0%; Score 53; DB 1; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.0059;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EGPWLEEEE 9

Db 1 EGPWLEEEE 9  
|||||

RESULT 10  
PCT-US94-13205-1  
; Sequence 1, Application PC/TUS9413205  
; GENERAL INFORMATION:  
; APPLICANT: Gevas, Phillip C.  
; APPLICANT: Grimes, Stephen  
; APPLICANT: Karr, Stephen  
; APPLICANT: Michaeli, Dov  
; APPLICANT: Schiensi, Robert  
; TITLE OF INVENTION: IMPROVED IMMUNOGENIC COMPOSITIONS  
; TITLE OF INVENTION: AGAINST HUMAN GASTRIN 17  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dimitrios T. Drivas, White and Case  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 100036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/13205  
; FILING DATE: 12-NOV-1993  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Drivas, Dimitrios T.  
; REGISTRATION NUMBER: 32,218  
; REFERENCE/DOCKET NUMBER: 1102865-028  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 819-8286  
; TELEFAX: (212) 354-8113  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 16 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; PCT-US94-13205-1

Query Match 100.0%; Score 53; DB 5; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.0059;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
|||||

RESULT 11  
US-09-174-216-4  
; Sequence 4, Application US/09174216A  
; Patent No. 6335176  
; GENERAL INFORMATION:  
; APPLICANT: Inglesse, James  
; APPLICANT: Glickman, Joseph Fraser  
; TITLE OF INVENTION: Incorporation of Phosphorylation Sites  
; FILE REFERENCE: 1073.050  
; CURRENT APPLICATION NUMBER: US/09/174.216A  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 4

; LENGTH: 17  
; TYPE: PPT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:Designed  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (1)  
; OTHER INFORMATION: PYRROLIDONE CARBOXYLIC ACID  
US-09-174-216-4

Query Match 100.0%; Score 53; DB 4; Length 17;  
Best Local Similarity 100.0%; Pred. No. 0.0062;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
|||||

Db 1 EGPWLEEEE 9  
|||||

RESULT 12  
US-08-446-692-75  
; Sequence 75, Application US/08446692  
; Patent No. 5759551  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Maria C.H. Lin  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,692  
; FILING DATE: 7-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maria C.H. Lin  
; REGISTRATION NUMBER: 29,323  
; REFERENCE/DOCKET NUMBER: 1151-4146 US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)415-8745  
; TELEFAX: (516)751-6849  
; INFORMATION FOR SEQ ID NO: 75:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-446-692-75

Query Match 94.3%; Score 50; DB 1; Length 12;  
Best Local Similarity 88.9%; Pred. No. 0.014;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
|||||

Db 1 QGPWLEEEE 9

RESULT 13

```
US-08-488-351A-75
; Sequence 75, Application US/08488351A
; Patent No. 5843446
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,351A
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/446,692
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/229,275
; FILING DATE: 14-APR-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057,166
; FILING DATE: 27-APR-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C.H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4146 US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)415-8745
; TELEFAX: (516)751-6849
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-488-351A-75

Query Match 94.3%; Score 50; DB 2; Length 12;
Best Local Similarity 88.8%; Pred. No. 0.014;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
Db 1 QGPWLEEEE 9

RESULT 14
US-08-488-692-74
; Sequence 74, Application US/08446692
; Patent No. 5759551
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,351A
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/446,692
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/229,275
; FILING DATE: 14-APR-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057,166
; FILING DATE: 27-APR-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C.H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4146 US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)415-8745
; TELEFAX: (516)751-6849
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-488-351A-75

Query Match 94.3%; Score 50; DB 2; Length 12;
Best Local Similarity 88.8%; Pred. No. 0.014;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
Db 1 QGPWLEEEE 9

RESULT 15
US-08-488-351A-74
; Sequence 74, Application US/08488351A
; Patent No. 5843446
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,351A
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/446,692
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/229,275
; FILING DATE: 14-APR-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057,166
; FILING DATE: 27-APR-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C.H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4146 US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)415-8745
; TELEFAX: (516)751-6849
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-446-692-74

Query Match 94.3%; Score 50; DB 1; Length 17;
Best Local Similarity 88.9%; Pred. No. 0.02;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
Db 1 QGPWLEEEE 9
```

FILING DATE: 14-APR-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/057,166  
FILING DATE: 27-APR-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maria C.H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4146 US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)415-8745  
TELEFAX: (516)751-6849  
INFORMATION FOR SEQ ID NO: 74:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-488-351A-74

Query Match 94.3%; Score 50; DB 2; Length 17;  
Best Local Similarity 88.9%; Pred. No. 0.02;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
Db 1 QGPWLEEEE 9

Search completed: February 5, 2003, 09:50:17  
Job time : 10.8571 secs





DR WPI; 1997-415075/38.  
 PT Treatment of glycine-extended gastrin-17-dependent gastrointestinal  
 PT tumours - using anti-G17 immunogenic composition, especially for  
 PT treatment of colorectal adenocarcinomas  
 XX  
 PS Example 1; Figure 1C; 37pp; English.  
 XX  
 CC The present sequence represents a specific example of an anti-gastrin-17  
 CC immunogen, comprising the N-terminal 9 amino acids of gastrin-17  
 CC conjugated to an immunogenic carrier such as Diphtheria toxoid, by  
 CC a spacer peptide. This immunogenic composition is used in a new treatment  
 CC of glycine-extended gastrin-17 (G17-Gly)-dependent gastrointestinal  
 CC tumours. Anti-G17 immunogens raise antibodies which bind both the  
 CC amidated and glycine-extended forms of G17. Neutralisation of progastrin  
 CC G17-Gly prohormone by the antibodies inhibits the growth of tumour cells  
 CC dependent on progastrin G17-Gly as growth stimulator or inducer. The  
 CC method is especially for the treatment of colorectal adenocarcinomas in  
 CC humans. The novel method is non-invasive, selectively reversible, does  
 CC not damage normal tissue, does not require frequent repeated treatments  
 CC and does not cross the blood brain barrier.  
 XX  
 SQ Sequence 9 AA;  
 Query Match 100.0%; Score 53; DB 18; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EGPWLEEEE 9  
 DB 1 EGPWLEEEE 9  
 RESULT 2  
 AAY51305  
 ID AAY51305 standard; peptide; 9 AA.  
 XX  
 AC AAY51305;  
 XX  
 DT 14-APR-2000 (first entry)  
 XX  
 DE Human gastrin aminoterminal (1-9) G17 peptide.  
 XX  
 KW Immunogenic; treatment; hypergastrinemia; gastrin G17; gastrin G34;  
 KW pernicious anemia; anti-ulcer; proton pump inhibitor; colorectal disease.  
 XX  
 OS Homo sapiens.  
 PN WO9959631-A1.  
 XX  
 PD 25-NOV-1999.  
 XX  
 PF 14-MAY-1999; 99WO-US10751.  
 XX  
 PR 15-MAY-1998; 98US-0085714.  
 XX  
 PA (APHT-) APHTON CORP.  
 XX  
 PI Gevas PC, Grimes S, Karr S, Michaeli D, Watson S;  
 XX  
 DR WPI; 2000-116301/10.  
 XX  
 PT Treating or preventing hypergastrinemia comprising administration of,  
 PT e.g. anti-gastrin antibodies -  
 XX  
 PS Disclosure; Page 11; 44pp; English.  
 XX  
 CC This invention describes a novel method for the treatment or preventing  
 CC hypergastrinemia by administering to a patient a gastrin G17 and/or G34  
 CC peptide fragment linked by an amino acid spacer to an immunogenic  
 CC carrier. The methods are used to treat hypergastrinemic patients,  
 CC particularly those with pernicious anemia, those receiving treatment  
 CC with anti-ulcer agents such as proton pump inhibitors (particularly

CC omeprazole or lansoprazole) or H<sub>2</sub> receptor blocking agents or  
 CC antagonists, or those having colorectal disorders or diseases. This  
 CC sequence represents the human gastrin G17 peptide which is used to  
 CC illustrate the method of the invention.  
 XX  
 SQ Sequence 9 AA;  
 Query Match 100.0%; Score 53; DB 21; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EGPWLEEEE 9  
 DB 1 EGPWLEEEE 9  
 RESULT 3  
 AAY59434  
 ID AAY59434 standard; peptide; 9 AA.  
 XX  
 AC AAY59434;  
 XX  
 DT 21-MAR-2000 (first entry)  
 XX  
 DE Antigastrin-G17 immunogen.  
 XX  
 KW Antigastrin-G17 immunogen; tumour; tumour growth factor;  
 KW gastrin-dependent tumour; therapy.  
 XX  
 OS Rattus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 1 /note= "pyroglutamic acid"  
 FT  
 XX  
 PN WO9959628-A2.  
 XX  
 PD 25-NOV-1999.  
 XX  
 PF 14-MAY-1999; 99WO-US10750.  
 XX  
 PR 15-MAY-1998; 98US-0085687.  
 XX  
 PA (APHT-) APHTON CORP.  
 XX  
 PI Gevas PC, Grimes S, Karr SL, Watson SA, Michaeli D;  
 XX  
 DR WPI; 2000-072406/06.  
 XX  
 PT Combination therapy, useful for treating tumours, especially -  
 XX  
 PS Claim 7; Page 23; 25pp; English.  
 XX  
 CC This sequence represents an antigastrin-G17 immunogen.  
 CC The invention relates to a method for treating tumours by immunologically  
 CC neutralising a tumour growth factor and administering one or more  
 CC chemotherapeutic agents. The method is useful for treating tumours,  
 CC especially gastrin-dependent tumours, including colorectal, stomach,  
 CC pancreatic and hepatocellular adenocarcinomas. The combination of  
 CC anti-G17 immunisation and chemotherapeutic agents increases the  
 CC therapeutic effects in controlling or inhibiting colorectal tumour growth  
 CC over chemotherapy alone.  
 XX  
 SQ Sequence 9 AA;  
 Query Match 100.0%; Score 53; DB 21; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EGPWLEEEE 9  
 DB 1 EGPWLEEEE 9

RESULT 4  
 AAR06245  
 ID AAR06245 standard; peptide; 12 AA;  
 XX  
 AC AAR06245;  
 DT 07-DEC-1990 (first entry)  
 XX  
 DE Antigenic peptide fragment selected from the 12 N-terminal AAs of  
 DE heptadecagastrin (G17).  
 XX  
 XX Gastrin; tumours; peptic ulcers; diptheria toxoid; tetanus toxin;  
 KW  
 XX  
 PN EP380230-A.  
 XX  
 PD 01-AUG-1990.  
 XX  
 PF 17-JAN-1990; 90EP-0300456.  
 XX  
 PR 12-MAY-1989; 89US-0351193.  
 PR 24-JAN-1989; 89US-0301353.  
 XX  
 PA (APHT-) APHTON CORP.  
 XX  
 PI Gevas PC, Grimes S, Karr SL, Littenberg RL;  
 XX  
 DR WPI; 1990-233029/31.  
 XX  
 XX Immunogens against gastrin peptide(s) - used to induce antibodies  
 PT that specifically neutralise single form of gastrin, G17 or G34  
 XX  
 PS Claim 6; Page 19; 32pp; English.  
 XX  
 CC Antigenic fragments may be attached to an immunogenic carrier  
 CC and used to raise Abs to a specific single form of Gastrin  
 CC ie, G17 or G34. Peptide fragments capable of binding to these  
 CC Abs are useful in neutralising anti-gastrin Abs in vivo.  
 XX  
 XX  
 SQ Sequence 12 AA;  
 Query Match 100.0%; Score 53; DB 11; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 0.023;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 EGPWLEEEE 9  
 Db 1 EGPWLEEEE 9  
 RESULT 5  
 AAR74297  
 ID AAR74297 standard; peptide; 15 AA;  
 XX  
 AC AAR74297;  
 XX  
 DT 10-JAN-1996 (first entry)  
 XX  
 DE Human gastrin 17 antigenic peptide hG17(1-9)-Arg9.  
 XX  
 KW Human gastrin 17; antigenic peptide hG17(1-9)-Arg9; immunisation;  
 KW treatment; gastro-oesophageal reflux disease; gastric; duodenal;  
 KW ulceration; cancer.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9513297-A2.  
 XX  
 PD 18-MAY-1995.  
 XX  
 PF 10-NOV-1994; 94WO-US13205.  
 XX  
 PR 12-NOV-1993; 93US-0151219.  
 XX  
 PA (APHT-) APHTON CORP.  
 XX  
 PI Gevas PC, Grimes S, Karr SL, Michaeli D, Scibienski R;  
 XX  
 DR WPI; 1995-194034/25.  
 XX  
 XX Immunogenic compsn. for producing anti-human gastrin 17 antibodies  
 PT - used for treating e.g. gastro-oesophageal reflux disease, gastric  
 PT and duodenal ulceration or cancer.  
 XX  
 XX Claim 1; Page 14; 17pp; English.  
 XX  
 PS AAR74295 and AAR74297 are human gastrin 17 (hG17) antigenic peptides,  
 CC used to produce anti-hG17 antibodies (Abs). The Abs can be  
 CC induced in a patient, or used for passive immunisation, for the  
 CC treatment of diseases in which hG17 is involved, e.g. gastric and  
 CC duodenal ulceration; gastro-oesophageal reflux disease and cancer.  
 CC  
 XX  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 53; DB 11; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 0.023;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 EGPWLEEEE 9  
 Db 1 EGPWLEEEE 9

XX  
 PA (APHT-) APHTON CORP.  
 XX  
 PI Gevas PC, Grimes S, Karr SL, Michaeli D, Scibienski R;  
 XX  
 DR WPI; 1995-194034/25.  
 XX  
 XX Immunogenic compsn. for producing anti-human gastrin 17 antibodies  
 PT - used for treating e.g. gastro-oesophageal reflux disease, gastric  
 PT and duodenal ulceration or cancer.  
 XX  
 XX Example 1; Page 3; 17pp; English.  
 XX  
 CC AAR74295 and AAR74297 are human gastrin 17 (hG17) antigenic peptides,  
 CC used to produce anti-hG17 antibodies (Abs). The Abs can be  
 CC induced in a patient, or used for passive immunisation, for the  
 CC treatment of diseases in which hG17 is involved, e.g. gastric and  
 CC duodenal ulceration; gastro-oesophageal reflux disease and cancer.  
 CC  
 XX  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 53; DB 16; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 0.029;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 EGPWLEEEE 9  
 Db 1 EGPWLEEEE 9  
 RESULT 6  
 AAR74295  
 ID AAR74295 standard; peptide; 16 AA.  
 XX  
 AC AAR74295;  
 XX  
 DT 10-JAN-1996 (first entry)  
 XX  
 DE Human gastrin 17 antigenic peptide hG17(1-9)-Ser9.  
 XX  
 KW Human gastrin 17; antigenic peptide hG17(1-9)-Ser9; immunisation;  
 KW treatment; gastro-oesophageal reflux disease; gastric; duodenal;  
 KW ulceration; cancer.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9513297-A2.  
 XX  
 PD 18-MAY-1995.  
 XX  
 PF 10-NOV-1994; 94WO-US13205.  
 XX  
 PR 12-NOV-1993; 93US-0151219.  
 XX  
 PA (APHT-) APHTON CORP.  
 XX  
 PI Gevas PC, Grimes S, Karr SL, Michaeli D, Scibienski R;  
 XX  
 DR WPI; 1995-194034/25.  
 XX  
 XX Immunogenic compsn. for producing anti-human gastrin 17 antibodies  
 PT - used for treating e.g. gastro-oesophageal reflux disease, gastric  
 PT and duodenal ulceration or cancer.  
 XX  
 XX Claim 1; Page 14; 17pp; English.  
 XX  
 PS AAR74295 and AAR74297 are human gastrin 17 (hG17) antigenic peptides,  
 CC used to produce anti-hG17 antibodies (Abs). The Abs can be  
 CC induced in a patient, or used for passive immunisation, for the  
 CC treatment of diseases in which hG17 is involved, e.g. gastric and  
 CC duodenal ulceration; gastro-oesophageal reflux disease and cancer.  
 CC  
 XX  
 SQ Sequence 16 AA;

Query Match 100.0%; Score 53; DB 16; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.031;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
| | | | | | | |  
Db 1 EGPWLEEEE 9

RESULT 7  
AA49309  
ID AAY49309 standard; peptide; 16 AA.  
XX  
AC AAY49309;  
XX  
DT 06-MAR-2000 (first entry)  
XX  
DE Human heptadecagastrin (G17) immunomimic with carboxy terminal spacer.

XX  
KW Gastroesophageal reflux disease; GERD; gastrin; anti-gastrin antibody;  
KW histamine H<sub>2</sub>; proton pump inhibitor; acid output; stomach; therapy;  
KW esophagitis; immunogen; human; heptadecagastrin; immunomimic; G17.  
XX  
OS Synthetic.  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers  
FT Modified-site 1  
FT /label= pGlu  
FT /note= "pyroglutamate"  
FT 10..16  
FT Peptide  
FT /note= "carboxy-terminal spacer"

XX  
PN WO9959612-A1.  
XX  
XX  
PD 25-NOV-1999.  
XX  
XX  
PF 14-MAY-1999; 99WO-US10734.  
XX  
XX  
PR 15-MAY-1998; 98US-0085610.  
XX

XX (APHT-) APHTON CORP.

XX Gevas PC, Grimes S, Karr S, Michaeli D;

XX WPI; 2000-062378/05.

XX Method for treatment of gastroesophageal reflux disease (GERD)

XX Example 5; Page 13; 24pp; English.

XX  
PS The invention relates to the treatment of gastroesophageal reflux disease  
CC (GERD) that comprises administering to a patient an immunogenic  
CC composition which generates anti-gastrin antibodies, which bind to  
CC gastrin, in a patient, and administering histamine H<sub>2</sub> antagonist or a  
CC proton pump inhibitor. The method provides a more effective method for  
CC controlling acid output by the stomach. The therapy is less costly. High  
CC gastrin levels associated with standard therapies are neutralized and  
CC undesirable side effects are reduced. The method permits a reduced dosage  
CC of acid reducing agent both at the acid producing level as well as the  
CC acid production stimulating level (gastrin). Reduction of dosages is  
CC desirable for prolonged treatment of GERD. In a combination therapy with  
CC H<sub>2</sub> agonists or proton pump inhibitors, anti-gastrin 17 antibody titers  
CC can be maintained by occasional booster shots while gastric acid  
CC inhibitor dosing is reduced. Immunization allows a sufficient time for  
CC the esophagitis to completely heal and no surgery is required. The  
CC present sequence represents a human heptadecagastrin (G17) immunomimic  
CC followed by a carboxy-terminal spacer.

XX Sequence 16 AA;

Query Match 100.0%; Score 53; DB 21; Length 16;

Best Local Similarity 100.0%; Pred. No. 0.031;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
| | | | | | | |  
Db 1 EGPWLEEEE 9

RESULT 8  
AAW65184  
ID AAW65184 standard; peptide; 17 AA.  
XX  
AC AAW65184;  
XX  
DT 02-OCT-1998 (first entry)  
XX  
DE Gastrin fragment analogue.  
XX  
KW Bradykinin; N-benzylglycine; agonist; receptor study; antagonist;  
KW achiral; analgesic; angiotensin II; gastrin.  
XX  
OS Synthetic.

XX  
FH Key Location/Qualifiers  
FT Modified-site 1  
FT /note= "pyroglutamic acid"  
FT Modified-site 17  
FT /note= "C-terminal amide"

XX  
PN US527882-A.  
XX  
PD 18-JUN-1996.  
XX  
XX  
PF 07-NOV-1994; 94US-0335202.  
XX  
XX  
PR 07-JUL-1989; 89US-0376839.  
PR 16-SEP-1992; 92US-0945664.  
PR 07-NOV-1994; 94US-0335202.  
XX  
XX (REGC ) UNIV CALIFORNIA.

XX Mitchell AR, Young JD;

XX WPI; 1996-299898/30.

XX New bradykinin analogues contg. N-benzyl-glycine - useful as  
XX bradykinin agonists or antagonists, useful e.g. as analgesics

XX Disclosure; Columns 7-8; 15pp; English.

XX  
CC The invention relates to the obtaining of a potent agonist or antagonist  
CC peptide by the replacement of selected amino acids with synthetic  
CC achiral amino acids. The present sequence represents a gastrin fragment  
CC analogue, where at least one of Tyr12 and Phe17 is intended to be  
CC replaced by N-benzylglycine. N-cyclohexylmethylglycine or the ring  
CC substituted derivatives thereof.

XX Sequence 17 AA;

Query Match 100.0%; Score 53; DB 17; Length 17;  
Best Local Similarity 100.0%; Pred. No. 0.033;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
| | | | | | | |  
Db 1 EGPWLEEEE 9

RESULT 9  
AAW24398  
ID AAW24398 standard; peptide; 17 AA.  
XX  
AC AAW24398;

```

XX 13-MAR-1998 (first entry)
XX Carboxy-amidated gastrin-17.
XX Carboxy-amidated gastrin-17; gastrointestinal tumour; immunogen;
KW colorectal adenocarcinoma; antibody; progastrin; cholecystochinin B.
KW anti-gastrin-17; anti-G-17.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Modified-site 17
XX /note= "C-terminal amide"
XX
XX WO9728821-A1.
XX
XX 14-AUG-1997.
XX
XX 07-FEB-1997; 97WO-US02029.
XX
XX 08-FEB-1996; 96US-0011411.
XX
XX (APHT-) APHTON CORP.
XX
XX Gevas PC, Grimes S, Karr SL, Michaeli D, Watson SA;
XX
XX WPI; 1997-415075/38.
XX
XX Treatment of glycine-extended gastrin-17-dependent gastrointestinal
XX tumours - using anti-G17 immunogenic composition, especially for
XX treatment of colorectal adenocarcinomas
XX
XX Example 1; Figure 1B; 37pp; English.
XX
XX The present sequence represents carboxy-amidated gastrin-17. Treatment
XX of gastrin-17-dependent gastrointestinal tumours comprises administering
XX to a mammal an anti-gastrin 17 (G17) immunogenic composition. Anti-G17
XX immunogens raise antibodies which bind both the amidated and glycine-
XX extended forms of G17. Neutralisation of progastrin G17-Gly prohormone
XX by the antibodies inhibits the growth of tumour cells dependent on
XX progastrin G17-Gly as growth stimulator or inducer. The method is
XX especially for the treatment of colorectal adenocarcinomas in humans.
XX This novel method is non-invasive, selectively reversible, does not
XX damage normal tissue, does not require frequent repeated treatments
XX and does not cross the blood brain barrier.
XX
XX Sequence 17 AA;
XX
XX Query Match 100.0%; Score 53; DB 18; Length 17;
XX Best Local Similarity 100.0%; Pred. No. 0.033;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 EGPWLEEEE 9
XX |
XX Db 1 EGPWLEEEE 9
XX
XX RESULT 10
XX AAU05580
XX ID AAU05580 standard; Peptide; 17 AA.
XX
XX AAU05580;
XX
XX 24-OCT-2001 (first entry)
XX
XX Human Janus kinase 3 (JAK3) biotinylated peptide substrate GAS1.
XX
XX Janus kinase 3; JAK3; JAK/STAT inhibitor; peptide substrate;
KW signal transducer and activator of transcription; osteoarthritis;
KW degenerative joint disease; rheumatoid arthritis; leprosy; asthma;
KW cancer; tumour; leukaemia.
XX

```

```

OS Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1
XX /note= "Glu is biotinylated"
XX Modified-site 17
XX /note= "Phe is amidated"
XX
XX WO200152892-A2.
XX
XX 26-JUL-2001.
XX
XX 22-JAN-2001; 2001WO-US02033.
XX
XX 24-JAN-2000; 2000US-0177872.
XX 28-NOV-2000; 2000US-0723490.
XX
XX (GENZ ) GENZYME CORP.
XX
XX Vasios G;
XX
XX WPI; 2001-465338/50.
XX
XX Use of inhibitors of Janus kinase/signal transducers and activators of
XX transcription for inhibiting onset and progression of degenerative
XX joint diseases or disorders such as osteoarthritis, rheumatoid
XX arthritis
XX
XX Example 6; Page 18; 55pp; English.
XX
XX The sequence represents a synthetic biotinylated peptide substrate for
XX human Janus kinase 3 (JAK3). The invention relates to the use of
XX JAK/STAT (Janus kinase/signal transducer and activator of
XX transcription) inhibitors other than debrromenaldesine (DBH) and
XX hymenaldesine (H) for inhibiting the progression or the likelihood of
XX developing diseases involving cartilage degradation, and for regulating
XX the expression of pro-inflammatory agents or cytokines in a chondrocyte,
XX and cartilage degrading enzymes in a cell. A JAK3/STAT inhibitor of the
XX invention is useful for inhibiting progression or likelihood of the
XX developing osteoarthritis or rheumatoid arthritis. The inhibitor is also
XX useful for treating other JAK/STAT-mediated diseases or disorders,
XX including T cell-mediated disorders, mast cell-mediated disorders,
XX type 2 (cytokine hypersensitivity) disorders, B cell lymphoma, and
XX myeloid diseases. T cell-mediated disorders include human T cell
XX leukaemia/lymphoma virus (HTLV)-1, Sazory's syndrome, C-abl
XX transformation, natural killer-like T cell lymphomas (NK-like tumours)
XX and graft-vs-host disease; cytokine hypersensitivity disorders include
XX Leishmaniasis, leprosy, allergy and viral infections; mast cell-mediated
XX disorders include hay fever, asthma, hives and anaphylaxis; and
XX leukaemias and lymphomas include acute lymphocytic and lymphoblastic
XX leukaemias, B cell lymphomas and leukaemias of myeloid origin. DBH and H
XX are useful as therapeutic agents in cancers in which JAK3 plays a role
XX in the initiation or progression of tumourigenesis.
XX
XX Sequence 17 AA;
XX
XX Query Match 100.0%; Score 53; DB 22; Length 17;
XX Best Local Similarity 100.0%; Pred. No. 0.033;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 EGPWLEEEE 9
XX |
XX Db 1 EGPWLEEEE 9
XX
XX RESULT 11
XX AAB91246
XX ID AAB91246 standard; Peptide; 17 AA.
XX
XX AAB91246;
XX
XX 22-JUN-2001 (first entry)
XX

```

DE Gastrin releasing peptide (GRP) SEQ ID NO:422.  
XX Protection; endogenous therapeutic peptide; peptidase; conjugation;  
KW blood component; modification; succinimidyl; maleimido group; amino;  
KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX WO20069900-A2.  
XX 23-NOV-2000.  
XX  
XX 17-MAY-2000; 2000WO-US13576.  
XX  
XX 17-MAY-1999; 99US-0134406.  
PR 10-SEP-1999; 99US-0153406.  
PR 15-OCT-1999; 99US-0159783.  
XX  
XX (CONJ-) CONJUCHEM INC.  
XX  
XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;  
XX WPI; 2001-112059/12.  
XX  
XX Modifying and attaching therapeutic peptides to albumin prevents  
PT peptidase degradation, useful for increasing length of in vivo activity  
PT  
XX  
XX Disclosure; Page 336; 73pp; English.  
XX  
XX The present invention describes a modified therapeutic peptide (I)  
CC comprising a therapeutically active amino acid region (III) and a  
CC reactive group (II) (e.g., succinimidyl and maleimido groups) attached to  
CC a less therapeutically active amino acid region (IV), which covalently  
CC bonds with amino/hydroxyl/thiol groups on blood components to form a  
CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.  
CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth  
CC factors and neurotransmitters, to protect them from peptidase activity  
CC in vivo for the treatment of various disorders. Endogenous therapeutic  
CC peptides are not suitable as drug candidates as they require frequent  
CC administration due to rapid degradation by peptidases in the body.  
CC Modifying and attaching therapeutic peptides to albumin prevents or  
CC reduces the action of peptidases to increase length of activity (half  
CC life) and specificity as bonding to large molecules decreases  
CC intracellular uptake and interference with physiological processes.  
CC AAB90829 to AAB92441 represent peptides which can be used in the  
CC exemplification of the present invention.  
XX  
XX Sequence 17 AA:  
Query Match 100.0% Score 53; DB 22; Length 17;  
Best Local Similarity 100.0% Pred. No. 0.033;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9  
RESULT 12  
ID AAB59273 standard; peptide; 17 AA.  
XX  
XX AAB59273;  
XX  
XX 27-MAR-2001 (first entry)  
DT  
XX  
XX KS2-peptide substrate.  
DE  
XX Phosphorylation; kinase; insulin.  
KW  
XX Unidentified.  
OS

XX WO200075167-A2.  
XX 14-DEC-2000.  
XX  
XX 09-JUN-2000; 2000WO-US16025.  
XX  
XX 09-JUN-1999; 99US-0138311.  
PR 10-JUN-1999; 99US-0138438.  
PR 08-JUL-1999; 99US-0349733.  
PR 28-APR-2000; 2000US-0200594.  
XX  
XX (LJLB-) LJL BIOSYSTEMS INC.  
XX  
XX Sportsman JR, Hoekstra MF, Lee SK, Cairns N, Kauvar LM;  
XX WPI; 2001-091201/10.  
XX  
XX Assay for detecting phosphorylation and dephosphorylation modification  
PT of proteins by contacting luminescence peptide with a binding partner  
PT and measuring change in luminescence polarization  
XX  
XX Claim 70; Page 70; 89pp; English.  
XX  
XX The present invention relates to detecting addition or removal of a  
CC phosphate group to or from a substrate. The method involves contacting  
CC a luminescent peptide with a binding partner that binds specifically  
CC to a phosphorylated peptide without regard to the particular amino acid  
CC sequence of the peptide. The method is useful for detecting  
CC phosphorylation and dephosphorylation modifications of proteins,  
CC including kinases and phosphatases. The methods can be used to study  
CC the kinase activity of different receptors e.g. the insulin receptor  
CC and to find agonists and antagonists of these receptors.  
XX  
XX Sequence 17 AA;  
Query Match 100.0% Score 53; DB 22; Length 17;  
Best Local Similarity 100.0% Pred. No. 0.033;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9  
RESULT 13  
AAU76504  
ID AAU76504 standard; peptide; 17 AA.  
XX  
XX AAU76504;  
XX  
XX 05-JUN-2002 (first entry)  
DT  
XX Protein kinase A (PKA) substrate #3.  
DE  
XX Protein kinase A; phosphorylation.  
KW  
XX Synthetic.  
OS  
XX  
XX Key Location/Qualifiers  
FH Modified-site 1  
FT /label= OTHER  
FT /note= "Pyroglutamic acid"  
XX  
XX US6335176-B1.  
PN  
XX 01-JAN-2002.  
PD  
XX 16-OCT-1998; 98US-0174216.  
PF  
XX 16-OCT-1998; 98US-0174216.  
PR  
XX (PHAR-) PHARMACOPEDIA INC.  
PA

XX Inglese J, Glickman JF;  
 XX WPI; 2002-194620/25.  
 XX Reagent for phosphorylating a compound, comprises a moiety that is  
 PT specifically reactive with reactive site chain of the compound, a  
 PT linking moiety and a peptide sequence comprising kinase substrate  
 XX Disclosure; Column 8; 26pp; English.  
 XX The invention relates to a reagent (I) for incorporation of a  
 CC phosphorylation site by reaction with a reactive side chain of a compound  
 CC to be phosphorylated. (I) comprises a structure A-B-C, where A is a  
 CC moiety that is specifically reactive with a reactive side chain, C is a  
 CC peptide sequence comprising kinase substrate, and B is a linking moiety  
 CC selected from any one of the 19 compounds given in the specification  
 CC e.g. N-gamma-maleimidobutyryloxy-succinamide ester. (I) is useful for  
 CC phosphorylation, by reacting (I) with a compound to be phosphorylated  
 CC (e.g. protein or polypeptide) and then phosphorylating the resulting  
 CC product using a protein kinase under phosphorylating conditions  
 CC comprising 32p-phosphate or 33p-phosphate. The phosphorylated compounds  
 CC are useful in assays such as drug discovery. The method is suitable for  
 CC radioactively phosphorylating already synthesised proteins, without the  
 CC need to carry out recombinant methods to incorporate an amino acid  
 CC sequence. The method is highly adaptable and can be used to phosphorylate  
 CC a broad variety of compounds that contain reactive site groups. (I)  
 CC avoids production of proteins having an inaccessible kinase substrate  
 CC sequence as can result from known recombinant methods. Introduction of  
 CC multiple phosphorylation sites in proteins is possible merely by  
 CC increasing the ratio of reagent to protein, and the method of  
 CC phosphorylating does not interfere with the protein's function or become  
 CC inaccessible as a result of protein folding. The method allows labeled  
 CC products to be obtained that have a higher specific activity that is  
 CC normally obtained with recombinant methods. The present sequence  
 CC represents a protein kinase A (PKA) peptide substrate used in examples  
 CC that demonstrate the method of the invention.  
 XX SQ Sequence 17 AA;  
 Query Match 100.0%; Score 53; DB 23; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 0.033;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EGPWLEEEE 9  
 Db | | | | | | | |  
 Db 1 EGPWLEEEE 9  
 RESULT 14  
 AAW24397  
 ID AAW24397 standard; peptide; 18 AA.  
 XX AC AAW24397;  
 XX 12-MAR-1998 (first entry)  
 DE Glycine-extended gastrin-17.  
 DE Glycine-extended gastrin-17; gastrointestinal tumour; immunogen;  
 KW colorectal adenocarcinoma; antibody; progastrin; cholecystokinin B;  
 KW anti-gastrin-17; anti-G17.  
 XX Homo sapiens.  
 XX WO9728821-A1.  
 XX 14-AUG-1997.  
 XX 07-FEB-1997; 97WO-US02029.  
 XX 08-FEB-1996; 96US-0011411.  
 XX

PA (APHT-) APHTON CORP.  
 XX Gevas PC, Grimes S, Karr SL, Michaeli D, Watson SA;  
 XX WPI; 1997-415075/38.  
 XX Treatment of glycine-extended gastrin-17-dependent gastrointestinal  
 PT tumours - using anti-G17 immunogenic composition, especially for  
 PT treatment of colorectal adenocarcinomas  
 XX Example 1; Figure 1A; 37pp; English.  
 XX The present sequence represents glycine-extended gastrin-17. Treatment  
 CC of glycine-extended gastrin-17 (G17-Gly)-dependent gastrointestinal  
 CC tumours comprises administering an anti-gastrin 17 (anti-G17)  
 CC immunogenic composition. Anti-G17 immunogens raise antibodies which bind  
 CC both the amidated and glycine-extended forms of G17. Neutralisation of  
 CC progastrin G17-Gly pro hormone by the antibodies inhibits the growth of  
 CC tumour cells dependent on progastrin G17-Gly as growth stimulator or  
 CC inducer. The method is especially for the treatment of colorectal  
 CC adenocarcinomas in humans. This novel method is non-invasive,  
 CC selectively reversible, does not damage normal tissue, does not require  
 CC frequent repeated treatments and does not cross the blood brain  
 CC barrier.  
 XX SQ Sequence 18 AA;  
 Query Match 100.0%; Score 53; DB 18; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 0.035;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EGPWLEEEE 9  
 Db | | | | | | | |  
 Db 1 EGPWLEEEE 9  
 RESULT 15  
 AAR62740  
 ID AAR62740 standard; peptide; 12 AA.  
 XX AC AAR62740;  
 XX 21-SEP-1995 (first entry)  
 DT Gastrin hapten.  
 DE Helper T cell epitope; universal immune stimulator; invasin; hapten;  
 KW gastrin; peptic ulcers; gastrin-stimulated tumours.  
 XX Homo sapiens.  
 XX WO9425060-A.  
 XX 10-NOV-1994.  
 XX 28-APR-1994; 94WO-US04832.  
 XX 27-APR-1993; 93US-0057166.  
 XX 14-APR-1994; 94US-0229275.  
 XX (LADD/) LADD A E.  
 XX (WANG/) WANG C Y.  
 XX (ZAMB/) ZAMB T.  
 XX Ladd AE, Wang CY, Zamb T;  
 XX WPI; 1994-357910/44.  
 XX Immunogenic luteinising hormone releasing hormone peptide(s) -  
 PT that suppress LHRH activity in males and females  
 XX Claims 27, 37; Page 96; 213pp; English.  
 XX

CC Synthetic immunogenic peptides are provided in which a universal immune  
CC stimulator is linked to a peptide or protein hapten containing B cell  
CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes  
CC potent immune responses to the coupled peptide or protein. The  
CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)  
CC which elicits an immune response to the coupled peptide in members of  
CC a heterogeneous population expressing diverse HLA phenotypes, and (B)  
CC an adjuvant peptide sequence from the invasin protein of Yersinia.  
CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the  
CC invasin and Th domains and between the immune stimulator and hapten  
CC components. When the hapten is IHRH, then optionally the invasin domain  
CC can be omitted from the immune stimulator component.  
CC The present sequence is an example of a gastrin hapten which can be  
CC bound to the immune stimulator to form a vaccine for treating  
CC peptic ulcer disease or gastrin-stimulated tumours.

XX  
SQ Sequence 12 AA;

Query Match 94.3%; Score 50; DB 15; Length 12;  
Best Local Similarity 88.9%; Pred. No. 0.071;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EGPWLEEEE 9  
      :|||||||  
Db 1 QGPWLEEEE 9

Search completed: February 5, 2003, 09:47:56  
Job time : 27.5714 secs

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:49:21 ; Search time 6.85714 Seconds  
(without alignments)  
29.097 Million cell updates/sec

Title: us-09-700-329-1  
Perfect score: 53  
Sequence: 1 EGPWLEEEE 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 129505 seqs, 22169297 residues

Total number of hits satisfying chosen parameters: 129505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	48	90.6	16	9	US-10-192-257-4
2	40	75.5	182	10	US-09-764-864-1205
3	38	71.7	62	9	US-10-021-811-16
4	38	71.7	179	9	US-10-021-811-4
5	38	71.7	221	9	US-10-021-811-14
6	38	71.7	317	9	US-09-956-993-2
7	38	71.7	369	9	US-09-956-993-4
8	38	71.7	369	10	US-09-838-561-8
9	38	71.7	369	10	US-09-816-760-8
10	37	69.8	50	10	US-09-757-049A-4
11	37	69.8	50	10	US-09-757-049A-5
12	37	69.8	50	10	US-09-757-049A-6
13	37	69.8	52	10	US-09-912-962-13
14	37	69.8	52	10	US-09-912-962-14
15	37	69.8	206	9	US-10-021-811-34
16	36	67.9	99	10	US-09-828-648-7
17	36	67.9	261	9	US-10-004-551-26
18	36	67.9	525	9	US-10-004-551-18
19	36	67.9	525	9	US-10-004-551-20

20	36	67.9	525	9	US-10-004-551-22	Sequence 22, Appl
21	36	67.9	525	9	US-10-004-551-24	Sequence 24, Appl
22	36	67.9	993	9	US-10-004-551-14	Sequence 14, Appl
23	36	67.9	994	9	US-10-004-551-16	Sequence 16, Appl
24	36	67.9	1596	9	US-09-909-567B-47	Sequence 47, Appl
25	35	66.0	84	10	US-09-864-761-47725	Sequence 47725, A
26	35	66.0	90	9	US-10-021-811-52	Sequence 52, Appl
27	35	66.0	115	9	US-10-008-118A-36	Sequence 36, Appl
28	35	66.0	115	10	US-09-443-704-36	Sequence 36, Appl
29	35	66.0	122	9	US-10-021-811-62	Sequence 62, Appl
30	35	66.0	126	9	US-10-021-811-24	Sequence 24, Appl
31	35	66.0	149	10	US-09-764-864-1171	Sequence 1171, Ap
32	35	66.0	264	9	US-10-008-118A-16	Sequence 16, Appl
33	35	66.0	264	10	US-09-443-704-16	Sequence 16, Appl
34	35	66.0	302	9	US-10-008-118A-18	Sequence 18, Appl
35	35	66.0	302	10	US-09-443-704-18	Sequence 18, Appl
36	35	66.0	323	9	US-10-021-811-26	Sequence 26, Appl
37	35	66.0	332	9	US-10-021-811-54	Sequence 54, Appl
38	35	66.0	529	9	US-09-712-363-175	Sequence 175, App
39	35	66.0	576	10	US-09-931-087A-1	Sequence 1, Appli
40	35	66.0	576	10	US-09-931-087A-21	Sequence 21, Appl
41	35	66.0	576	10	US-09-931-087A-22	Sequence 22, Appl
42	35	66.0	576	10	US-09-931-087A-23	Sequence 23, Appl
43	34	64.2	94	9	US-10-021-811-28	Sequence 28, Appl
44	34	64.2	103	10	US-09-751-100B-76	Sequence 76, Appl
45	34	64.2	109	10	US-09-751-100B-96	Sequence 96, Appl

ALIGNMENTS

RESULT 1  
US-10-192-257-4  
; Sequence 4, Application US/10192257  
; Publication No. US20030021786A1  
; GENERAL INFORMATION:  
; APPLICANT: Aphton Corporation  
; TITLE OF INVENTION: Treatment and Prevention of Cancerous and Pre-Cancerous Condit  
; FILE REFERENCE: 1102865-0057  
; CURRENT APPLICATION NUMBER: US/10/192,257  
; CURRENT FILING DATE: 2002-07-09  
; PRIOR APPLICATION NUMBER: US 60/303,868  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: xaa=pyroglutamic acid  
US-10-192-257-4

Query Match 90.6%; Score 48; DB 9; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.038;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPWLEEEE 9  
Db 2 GPWLEEEE 9

RESULT 2  
US-09-764-864-1205  
; Sequence 1205, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PTZ23

```
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1205
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (23)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (108)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (129)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (177)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (182)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1205

Query Match          75.5%; Score 40; DB 10; Length 182;
Best Local Similarity 85.7%; Pred. No. 8.2;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PWLEEEE 9
Db 164 PWEDEE 170

RESULT 3
US-10-021-811-16
; Sequence 16, Application US/10021811
; Publication No. US20030024007A1
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Fang, Yiwen
; APPLICANT: Odell, Joan
; APPLICANT: Weng, Zude
; TITLE OF INVENTION: Plant Myb Transcription Factor Homologs
; FILE REFERENCE: BB1294 US NA
; CURRENT APPLICATION NUMBER: US/10/021,811
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/110,609
; PRIOR FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Oryza sativa
; NAME/KEY: UNSURE
; LOCATION: (59)
US-10-021-811-16

Query Match          71.7%; Score 38; DB 9; Length 62;
Best Local Similarity 55.6%; Pred. No. 6;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9
Db 10 KGPWTEQED 18

RESULT 4
US-10-021-811-4
; Sequence 16, Application US/10021811
; Publication No. US20030024007A1
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Fang, Yiwen
; APPLICANT: Odell, Joan
; APPLICANT: Weng, Zude
; TITLE OF INVENTION: Plant Myb Transcription Factor Homologs
; FILE REFERENCE: BB1294 US NA
; CURRENT APPLICATION NUMBER: US/10/021,811
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/110,609
; PRIOR FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Oryza sativa
; NAME/KEY: UNSURE
; LOCATION: (59)
US-10-021-811-4

Query Match          71.7%; Score 38; DB 9; Length 221;
Best Local Similarity 55.6%; Pred. No. 21;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9
Db 3 KGPWTEQED 11

RESULT 6
US-09-956-993-2
; Sequence 2, Application US/09956993
; Patent No. US20020164733A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME, NUCLEIC ACID
; FILE REFERENCE: CL001181DIV
; CURRENT APPLICATION NUMBER: US/09/956,993
; CURRENT FILING DATE: 2001-09-21
```

```
; Sequence 4, Application US/10021811
; Publication No. US20030024007A1
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Fang, Yiwen
; APPLICANT: Odell, Joan
; APPLICANT: Weng, Zude
; TITLE OF INVENTION: Plant Myb Transcription Factor Homologs
; FILE REFERENCE: BB1294 US NA
; CURRENT APPLICATION NUMBER: US/10/021,811
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/110,609
; PRIOR FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Zea mays
US-10-021-811-4

Query Match          71.7%; Score 38; DB 9; Length 179;
Best Local Similarity 55.6%; Pred. No. 17;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9
Db 10 KGPWTEQED 18

RESULT 5
US-10-021-811-14
; Sequence 14, Application US/10021811
; Publication No. US20030024007A1
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Fang, Yiwen
; APPLICANT: Odell, Joan
; APPLICANT: Weng, Zude
; TITLE OF INVENTION: Plant Myb Transcription Factor Homologs
; FILE REFERENCE: BB1294 US NA
; CURRENT APPLICATION NUMBER: US/10/021,811
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/110,609
; PRIOR FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-021-811-14

Query Match          71.7%; Score 38; DB 9; Length 221;
Best Local Similarity 55.6%; Pred. No. 21;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9
Db 3 KGPWTEQED 11

RESULT 6
US-09-956-993-2
; Sequence 2, Application US/09956993
; Patent No. US20020164733A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME, NUCLEIC ACID
; FILE REFERENCE: CL001181DIV
; CURRENT APPLICATION NUMBER: US/09/956,993
; CURRENT FILING DATE: 2001-09-21
```

; PRIOR APPLICATION NUMBER: 09/816,088  
; PRIOR FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 317  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-956-993-2

Query Match 71.7%; Score 38; DB 9; Length 317;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPWLEE 7  
Db 47 GPWLEE 52

RESULT 7  
US-09-956-993-4  
; Sequence 4, Application US/09956993  
; Patent No. US20020164733A1  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Ming-Hui et al  
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME, NUCLEIC ACID  
; FILE REFERENCE: CLO01181DIV  
; CURRENT APPLICATION NUMBER: US/09/956,993  
; PRIOR FILING DATE: 2001-09-21  
; PRIOR APPLICATION NUMBER: 09/816,088  
; PRIOR FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-956-993-4

Query Match 71.7%; Score 38; DB 9; Length 369;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPWLEE 7  
Db 47 GPWLEE 52

RESULT 8  
US-09-838-561-8  
; Sequence 8, Application US/09838561  
; Patent No. US20020042371A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; APPLICANT: Cook, William James  
; APPLICANT: Williamson, Mark  
; APPLICANT: Rudolph-Owen, Laura A.  
; APPLICANT: Gimeno, Ruth  
; TITLE OF INVENTION: 32142, 21481, 25964, 21686, NOVEL DEHYDROGENASE  
; FILE REFERENCE: MNI-134CP2  
; CURRENT APPLICATION NUMBER: US/09/838,561  
; PRIOR FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 09/816,760  
; PRIOR FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: 09/634,955  
; PRIOR FILING DATE: 2000-08-08  
; PRIOR APPLICATION NUMBER: 60/192,002  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8

; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-838-561-8

Query Match 71.7%; Score 38; DB 10; Length 369;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPWLEE 7  
Db 47 GPWLEE 52

RESULT 9  
US-09-816-760-8  
; Sequence 8, Application US/09816760  
; Patent No. US20020052032A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; APPLICANT: Cook, William James  
; APPLICANT: Williamson, Mark  
; APPLICANT: Rudolph-Owen, Laura A.  
; TITLE OF INVENTION: 32142, 21481, 25964, 21686, NOVEL HUMAN DEHYDROGENASE  
; FILE REFERENCE: MNI-134CP  
; CURRENT APPLICATION NUMBER: US/09/816,760  
; CURRENT FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: 09/634,955  
; PRIOR FILING DATE: 2000-08-08  
; PRIOR APPLICATION NUMBER: 60/192,002  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-816-760-8

Query Match 71.7%; Score 38; DB 10; Length 369;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPWLEE 7  
Db 47 GPWLEE 52

RESULT 10  
US-09-757-049A-4  
; Sequence 4, Application US/09757049A  
; Patent No. US20020127702A1  
; GENERAL INFORMATION:  
; APPLICANT: BERNSTEIN, Harold S.  
; APPLICANT: COUGHLIN, Shaun R.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR REGULATING CELL CYCLE  
; FILE REFERENCE: UCSF-020/02US  
; CURRENT APPLICATION NUMBER: US/09/757,049A  
; CURRENT FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/156,316  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: US 60/060,688  
; PRIOR FILING DATE: 1997-09-22  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 50  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-757-049A-4

Query Match 69.8%; Score 37; DB 10; Length 50;  
Best Local Similarity 55.6%; Pred. No. 7;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
:||||:|:  
Db 2 KGPWTKEED 10

## RESULT 11

US-09-757-049A-5  
; Sequence 5, Application US/09757049A  
; Patent No. US20020127702A1  
; GENERAL INFORMATION:  
; APPLICANT: BERNSTEIN, Harold S.  
; APPLICANT: COUGHLIN, Shaun R.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR REGULATING CELL CYCLE  
; FILE REFERENCE: UCSF-020/020S  
; CURRENT APPLICATION NUMBER: US/09/757,049A  
; CURRENT FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/156,316  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: US 60/060,688  
; PRIOR FILING DATE: 1997-09-22  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 50  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-757-049A-5

Query Match 69.8%; Score 37; DB 10; Length 50;  
Best Local Similarity 55.6%; Pred. No. 7;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
:||||:|:  
Db 2 KGPWTKEED 10

## RESULT 12

US-09-757-049A-6  
; Sequence 6, Application US/09757049A  
; Patent No. US20020127702A1  
; GENERAL INFORMATION:  
; APPLICANT: BERNSTEIN, Harold S.  
; APPLICANT: COUGHLIN, Shaun R.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR REGULATING CELL CYCLE  
; FILE REFERENCE: UCSF-020/020S  
; CURRENT APPLICATION NUMBER: US/09/757,049A  
; CURRENT FILING DATE: 2001-01-08  
; PRIOR APPLICATION NUMBER: US 09/156,316  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: US 60/060,688  
; PRIOR FILING DATE: 1997-09-22  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 50  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-757-049A-6

Query Match 69.8%; Score 37; DB 10; Length 50;  
Best Local Similarity 55.6%; Pred. No. 7;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
:||||:|:  
Db 2 KGPWTKEED 10

## RESULT 13

US-09-912-962-13  
; Sequence 13, Application US/09912962  
; Patent No. US20020076719A1  
; GENERAL INFORMATION:  
; APPLICANT: de Lange, Titia  
; APPLICANT: Broccoli, Dominique  
; APPLICANT: Smogorzewska, Agata  
; TITLE OF INVENTION: TELOMERE REPEAT BINDING FACTOR AND  
; DIAGNOSTIC AND THERAPEUTIC USE THEREOF  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KLAUBER & JACKSON  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/912,962  
; FILING DATE: 25-Jul-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/018,635  
; FILING DATE: 04-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: David A. Jackson  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/POCKET NUMBER: 600-1-142 CIP1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 52 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-09-912-962-13

Query Match 69.8%; Score 37; DB 10; Length 52;  
Best Local Similarity 55.6%; Pred. No. 7.3;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
:||||:|:  
Db 3 KGPWTKEED 11

## RESULT 14

US-09-912-962-14  
; Sequence 14, Application US/09912962  
; Patent No. US20020076719A1  
; GENERAL INFORMATION:  
; APPLICANT: de Lange, Titia  
; APPLICANT: Broccoli, Dominique  
; APPLICANT: Smogorzewska, Agata  
; TITLE OF INVENTION: TELOMERE REPEAT BINDING FACTOR AND  
; DIAGNOSTIC AND THERAPEUTIC USE THEREOF  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KLAUBER & JACKSON  
; STREET: 411 Hackensack Avenue

CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601

Search completed: February 5, 2003, 09:56:01  
Job time : 6.85714 secs

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/912,962  
FILING DATE: 25-Jul-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/018,635  
FILING DATE: 04-FEB-1998

ATTORNEY/AGENT INFORMATION:  
NAME: David A. Jackson  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-142 CIP1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521

INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 52 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-09-912-962-14

Query Match 69.8%; Score 37; DB 10; Length 52;  
Best Local Similarity 55.6%; Pred. No. 7.3;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
DB 3 KGPWTREED 11

## RESULT 15

US-10-021-811-34  
; Sequence 34, Application US/10021811  
; Publication No. US20030024007A1  
; GENERAL INFORMATION:  
; APPLICANT: Fahg, Yiwen  
; APPLICANT: Odell, Joan  
; APPLICANT: Weng, Zude  
; TITLE OF INVENTION: Plant Myb Transcription Factor Homologs  
; FILE REFERENCE: BBI294 US NA  
; CURRENT APPLICATION NUMBER: US/10/021,811  
; CURRENT FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/110,609  
; PRIOR FILING DATE: 1998-December-02  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: Microsoft office 97  
; SEQ ID NO 34  
; LENGTH: 206  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-10-021-811-34

Query Match 69.8%; Score 37; DB 9; Length 206;  
Best Local Similarity 55.6%; Pred. No. 28;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
DB 16 KGPWINEED 24



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 06:46:31 ; Search time 11.5714 Seconds  
(without alignments)  
74.771 Million cell updates/sec

Title: US-09-700-329-1

Perfect score: 53

Sequence: 1 EGPWLEEEE 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	94.3	33	2 A60506	big gastrin - North
2	50	94.3	101	1 GMHUB	gastrin precursor
3	50	94.3	104	1 GMCT	gastrin precursor
4	48	90.6	17	2 A60071	gastrin - rhesus m
5	48	90.6	104	1 GMPGB	gastrin precursor
6	48	90.6	104	1 GMDG	gastrin precursor
7	47	88.7	17	1 GMSH	gastrin - sheep
8	47	88.7	34	2 JS0426	big gastrin - goat
9	47	88.7	104	1 GMB0	gastrin precursor
10	45	84.9	238	2 T46166	MYB27 protein - Ar
11	43	81.1	16	2 A29541	little gastrin - C
12	43	81.1	335	2 C75119	deoxyhypusine synt
13	42	79.2	370	2 F64401	deoxyhypusine synt
14	41	77.4	248	2 C91080	hypothetical prote
15	41	77.4	248	2 D85925	hypothetical prote
16	41	77.4	273	1 JQ2390	MYB transcription
17	41	77.4	293	2 T09758	myb-related protei
18	40	75.5	33	1 GMPGB	big gastrin [valid
19	40	75.5	33	2 B29541	big gastrin - Chin
20	40	75.5	110	2 S20350	nepin nla - rape
21	40	75.5	268	1 JQ0961	myb-related protei
22	40	75.5	745	2 A71448	probable MYB trans
23	39	73.6	204	2 C35878	class I major hist
24	39	73.6	377	2 A49885	MHC class I histoc
25	39	73.6	379	2 E35878	class I major hist
26	39	73.6	379	2 A35878	class I major hist
27	39	73.6	406	2 B35878	class I major hist
28	39	73.6	498	2 S11246	LAG-3 protein prec
29	39	73.6	960	2 T00808	hypothetical prote

30 38 71.7 234 2 T51666 myb-related transc  
31 38 71.7 256 2 T49254 Myb DNA binding pr  
32 38 71.7 278 2 T03850 myb-related protei  
33 38 71.7 325 2 T51509 probable transcrip  
34 38 71.7 378 2 D82158 N-acetylglucosamin  
35 38 71.7 529 2 T48253 myb-like protein -  
36 38 71.7 894 2 T26149 hypothetical prote  
37 38 71.7 910 2 A48403 alpha-actinin - Ca  
38 38 71.7 920 2 T26147 hypothetical prote  
39 38 71.7 985 2 T41135 hypothetical prote  
40 38 71.7 4006 2 T09070 probable tenascin  
41 37 69.8 106 2 S26636 napin nlb - rape  
42 37 69.8 189 1 G69355 Mj0653 homolog AFO  
43 37 69.8 222 2 F69335 conserved hypochet  
44 37 69.8 301 1 A37766 SEC14 protein - ye  
45 37 69.8 301 2 T45859 R2R3-MYB transcrip

#### ALIGNMENTS

##### RESULT 1

A60506  
big gastrin - North American opossum  
N:Contains: gastrin  
C:Species: Didelphis virginiana, Didelphis marsupialis virginiana (North American opo  
C>Date: 19-Mar-1993 #sequence\_revision 19-Mar-1993 #text\_change 11-May-2000  
C:Accession: A60506  
R:Shinomura, Y.; Eng, J.; Rattan, S.C.; Yalow, R.S.  
Comp. Biochem. Physiol. B 96, 239-242, 1990  
A:Title: Opossum (Didelphis virginiana) "little" and "big" gastrins.  
A:Reference number: A60506; MUID:90298616; PMID:2361360  
A:Accession: A60506  
A:Molecule type: protein  
A:Residues: 1-93 <SHI>  
C:Superfamily: gastrin  
C:Keywords: amidated carboxyl end; hormone; pyroglutamic acid; sulfoprotein  
F:1-33/Product: big gastrin #status experimental <MAYB>  
F:18-33/Product: gastrin #status experimental <MAYB>  
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
F:28/Binding site: sulfate (Tyr) (covalent) (partial) #status experimental  
F:23/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 94.3% Score 50; DB 2; Length 33;  
Best Local Similarity 88.9%; Pred. No. 0.037;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9

Db 18 QGPWLEEEE 26

##### RESULT 2

GMHUB  
gastrin precursor [validated] - human  
N:Contains: big gastrin; cryptagastrin; gastrin; gastrin-17  
C:Species: Homo sapiens (man)  
C>Date: 24-Apr-1984 #sequence\_revision 30-Jun-1987 #text\_change 19-Jan-2001  
C:Accession: A93497; A94473; A93152; A18854; A40869; A32487; B32487;  
R:Ito, R.; Sato, K.; Helmer, T.; Jay, G.; Agarwal, K.  
Proc. Natl. Acad. Sci. U.S.A. 81, 4662-4666, 1984  
A:Title: Structural analysis of the gene encoding human gastrin: the large intron con  
A:Reference number: A93997; MUID:84272693; PMID:6087340  
A:Accession: A93997  
A:Molecule type: DNA  
A:Residues: 1-101 <ITPO>  
A:Cross-references: GB:K01254; GB:J00147; NID:q182987; PID:AA859533.1; PID:q182989  
R:Kato, K.; Hayashizaki, Y.; Takahashi, Y.; Himeno, S.; Matsubara, K.  
Nucleic Acids Res. 11, 8197-8203, 1983  
A:Title: Molecular cloning of the human gastrin gene.  
A:Reference number: A93497; MUID:84169471; PMID:6324077  
A:Accession: A93497  
A:Molecule type: DNA

A:Residues: 1-101 <KAT>  
A:Cross-references: GB:X00183; NID:g31648; PIDN:CAA25005.1; PID:g31649  
R:Harris, J.I.; Kenner, E.W.  
A:Title: unpublished results, cited by Gregory, R.A., and Tracy, H.J., in Gastrointestinal Hormones  
A:Reference number: A94473  
A:Accession: A94473  
A:Molecule type: protein  
A:Residues: 59-64, 'H', '66', 'S', '68-92 <H>  
A:Experimental source: gastrinoma tissue  
R:Bentley, P.H.; Kenner, G.W.; Sheppard, R.C.  
Nature 209, 583-585, 1966  
A:Title: Human gastrin isolation, structure and synthesis.  
A:Reference number: A93152; MUID:67021327; PMID:5921183  
A:Accession: A93152  
A:Molecule type: protein  
A:Residues: 76-92 <BEN>  
A:Note: gastrin comprises the carboxyl-terminal 17 residues of big gastrin  
R:Gregory, R.A.; Tracy, H.J.; Agarwal, K.L.; Grossman, M.I.  
Gut 10, 603-608, 1969  
A:Title: Amino acid constitution of two gastrins isolated from Zollinger-Ellison tumour  
A:Reference number: A91628; MUID:69298172; PMID:5822140  
A:Accession: A91628  
A:Molecule type: protein  
A:Residues: 76-92 <GRE>  
A:Note: the gastrin was isolated from human Zollinger-Ellison pancreatic tumor  
R:Wiborg, O.; Berglund, L.; Boel, E.; Norris, F.; Norris, K.; Rehfeld, J.F.; Marcker, K.  
Proc. Natl. Acad. Sci. U.S.A. 81, 1067-1069, 1984  
A:Title: Structure of a human gastrin gene.  
A:Reference number: A18854; MUID:84144842; PMID:6322186  
A:Accession: A18854  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-101 <WIB>  
A:Cross-references: GB:K01254; GB:J00147; NID:g182987; PIDN:AAB59533.1; PID:g182989  
R:Huebner, V.B.; Jiang, R.; Lee, T.D.; Legesse, K.; Walsh, J.H.; Shively, J.E.; Chew, P.  
J. Biol. Chem. 266, 12223-12227, 1991  
A:Title: Purification and structural characterization of progastrin-derived peptides from  
A:Reference number: A40869; MUID:91286236; PMID:2061307  
A:Accession: A40869  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 22-51 <HUE>  
R:Higashimoto, Y.; Himeno, S.; Shinomura, Y.; Nagao, K.; Tamura, T.; Tarui, S.  
Biochem. Biophys. Res. Commun. 160, 1364-1370, 1989  
A:Title: Purification and structural determination of urinary NH-2-terminal big gastrin  
A:Reference number: A32487; MUID:89273602; PMID:2730647  
A:Accession: A32487  
A:Molecule type: protein  
A:Residues: 59-67 <HI2>  
A:Experimental source: urine  
A:Note: this urinary fragment of big gastrin was designated peak Ia  
A:Accession: B32487  
A:Molecule type: protein  
A:Residues: 59-66 <HI3>  
A:Experimental source: urine  
A:Note: this urinary fragment of big gastrin was designated peak Ib  
A:Accession: C32487  
A:Molecule type: protein  
A:Residues: 59-68 <HI4>  
A:Experimental source: urine  
A:Note: this urinary fragment of big gastrin was designated peak II  
R:Higashimoto, Y.; Shinomura, Y.; Nagao, K.; Yasunaga, Y.; Tamura, T.; Tarui, S.  
Biochem. Biophys. Res. Commun. 172, 1392-1399, 1990  
A:Title: Purification of N-terminal hexapeptide of big gastrin from human urine.  
A:Reference number: A36249; MUID:91058586; PMID:2244919  
A:Accession: A36249  
A:Molecule type: protein  
A:Residues: 59-64 <HIG>  
A:Note: this urinary fragment of big gastrin was designated peak III  
R:Boel, E.; Vuust, J.; Norris, F.; Norris, K.; Wind, A.; Rehfeld, J.F.; Marcker, K.A.  
Proc. Natl. Acad. Sci. U.S.A. 80, 2866-2869, 1983  
A:Title: Molecular cloning of human gastrin cDNA: evidence for evolution of gastrin by g  
A:Reference number: I37408; MUID:83221503; PMID:6574456

A:Accession: I37408  
A:Molecule type: mRNA  
A:Residues: 1-101 <RES>  
A:Cross-references: EMBL:V00511; NID:g31654; PIDN:CAA23769.1; PID:g31655  
R:Rehfeld, J.F.; Johnsen, A.H.  
Eur. J. Biochem. 223, 765-773, 1994  
A:Title: Identification of gastrin component I as gastrin-71. The largest possible bi  
A:Reference number: S48183; MUID:94333379; PMID:8055952  
A:Accession: S48183  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 22-40 <REH>  
R:Kariya, Y.; Kato, K.H.; Hayashizaki, Y.; Himeno, S.; Tarui, S.; Matsubara, K.  
Gene 50, 345-352, 1986  
A:Title: Expression of human gastrin gene in normal and gastrinoma tissues.  
A:Reference number: I54006; MUID:87219893; PMID:3034736  
A:Accession: I54006  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-101 <RAR>  
A:Cross-references: GB:M15958; NID:g182990; PIDN:AAA52520.1; PID:g182991  
R:Rehfeld, J.F.; Hansen, C.P.; Johnsen, A.H.  
EMBO J. 14, 389-396, 1995  
A:Title: Post-poly(Glu) cleavage and degradation modified by O-sulfated tyrosine: a n  
A:Reference number: S54350; MUID:95137019; PMID:7530658  
A:Accession: S54350  
A:Molecule type: protein  
A:Residues: 76-92 <REW>  
A:Note: gastrin-6 isolated from gastric mucosa and blood is fully active; sulfation o  
C:Genetics  
A:Gene: GDB:GAS  
A:Cross-references: GDB:119261; OMIM:137250  
A:Map position: 17q-17q  
A:Introns: 71/1  
C:Superfamily: gastrin  
C:Keywords: amidated carboxyl end; phosphoprotein; pyroglutamic acid; sulfoprotein  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-56/Product: cryptogastrin (amino-terminal propeptide) #status experimental <PRO>  
F:59-92/Product: big gastrin #status experimental <BGN>  
F:76-92/Product: gastrin #status experimental <SGN>  
F:87-92/Product: gastrin-6 #status experimental <GN6>  
F:59/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #link BGN #sta  
F:76/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #link SGN #sta  
F:87/Binding site: sulfate (Tyr) (covalent) (partial) #status experimental  
F:92/Modified site: amidated carboxyl end (Phe) (amide in mature form from following  
Query Match 94.3% Score 50; DB 1; Length 101;  
Best Local Similarity 88.9%; Pred. No. 0.13;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 EGPWLEEEE 9  
:|||||||  
Db 76 QGPWLEEEE 84  
RESULT 3  
GMCT  
gastrin precursor [validated] - cat  
N:Contains: big gastrin (gastrin-34); gastrin  
C:Species: Felis silvestris catus (domestic cat)  
C:Date: 13-Jun-1983 #sequence, revision 02-Jun-1994 #text\_change 20-Oct-2000  
C:Accession: S14401; A01621; A61074  
R:Kim, S.J.; Uhm, K.N.; Kang, Y.K.; Yoo, O.J.  
DNA Seq. 1, 181-187, 1991  
A:Title: Bovine and feline gastrin cDNA sequences and the amino acid and nucleotide s  
A:Reference number: S14400; MUID:92127058; PMID:1773057  
A:Accession: S14401  
A:Molecule type: mRNA  
A:Residues: 1-104 <KIT>  
A:Cross-references: EMBL:X16582; NID:g1099; PIDN:CAA34599.1; PID:g1100  
R:Agarwal, K.L.; Kenner, G.W.; Sheppard, R.C.  
J. Am. Chem. Soc. 91, 3096-3097, 1969  
A:Title: Feline gastrin. An example of peptide sequence analysis by mass spectrometry

A:Reference number: A01621; MUID:69206035; PMID:5784957

A:Accession: A01621

A:Molecule type: protein

A:Residues: 76-92 <AGA>

R:Eng, J.; Du, B.H.; Johnson, G.F.; Kapakamedala, S.; Samuel, S.; Raufman, J.P.; Straus, Regul. Pept. 37, 9-13, 1992

A:Title: Cat gastrinoma and the sequence of cat gastrins.

A:Reference number: A61074; MUID:92262853; PMID:1585019

A:Accession: A61074

A:Molecule type: protein

A:Residues: 59-92 <ENG>

C:Superfamily: gastrin

C:Keywords: amidated carboxyl end; phosphoprotein; pyroglutamic acid; sulfoprotein

F:1-19/Domain: signal sequence; #status predicted <SIG>

F:59-92/Product: big gastrin #status experimental <BMAT>

F:76-92/Product: gastrin #status experimental <WAT>

F:76/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental

F:12/Binding site: sulfate (Tyr) (covalent) (partial) #status experimental

F:92/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gly

Query Match 94.3%; Score 50; DB 1; Length 104;

Best Local Similarity 88.9%; Pred. No. 0.14;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEE 9

Db 76 QGPWLEEE 84

#### RESULT 4

A60071

gastrin - rhesus macaque

C:Species: Macaca mulatta (rhesus macaque)

C:Date: 01-Dec-1992 #sequence\_revision 01-Dec-1992 #text\_change 20-Mar-1998

A:Reference number: A60071; MUID:91164506; PMID:2003150

A:Accession: A60071

A:Molecule type: protein

A:Residues: 1-17 <YUA>

C:Superfamily: gastrin

C:Keywords: amidated carboxyl end; hormone; intestine; pyroglutamic acid; sulfoprotein

F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:12/Binding site: sulfate (Tyr) (covalent) (partial) #status experimental

F:17/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 90.6%; Score 48; DB 2; Length 17;

Best Local Similarity 77.8%; Pred. No. 0.039;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEE 9

Db 1 QGPWLEEE 9

#### RESULT 5

GMGB

gastrin precursor [validated] - pig

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 17-Dec-1982 #sequence\_revision 17-Dec-1982 #text\_change 31-Dec-2000

C:Accession: A93903; B94473; A93148; I46622; A60070; A01618

R:Yoo, O.J.; Powell, C.T.; Agarwal, K.L.

Proc. Natl. Acad. Sci. U.S.A. 79, 1049-1053, 1982

A:Title: Molecular cloning and nucleotide sequence of full-length cDNA coding for porcine

A:Reference number: A93903; MUID:82174533; PMID:6951161

A:Accession: A93903

A:Molecule type: mRNA

A:Residues: 1-104 <YOO>

A:Cross-references: GB:V01303; GB:J00651; NID:g1846; PIDN:CAA24610.1; PID:g1847

R:Harris, J.I.; Kenner, E.W.

unpublished results, cited by Gregory, R.A., and Tracy, H.J., in Gastrointestinal Hormon

A:Reference number: A94473

A:Accession: B94473

A:Molecule type: protein

A:Residues: 59-64, Hpp, 68-92 <HAR>

A:Note: This peptide was sulfated in two-thirds of the molecules

R:Gregory, H.; Hardy, P.M.; Jones, D.S.; Kenner, G.W.; Sheppard, R.C.

Nature 204, 931-933, 1964

A:Title: The antral hormone gastrin.

A:Reference number: A93148

A:Accession: A93148

A:Molecule type: protein

A:Residues: 76-92 <GR>

R:Anderson, J.C.; Barton, M.A.; Gregory, R.A.; Hardy, P.M.; Kenner, G.W.; MacLeod, J.

Nature 204, 933-934, 1964

A:Reference number: A93149

A:Contents: annotation; synthesis

R:Agarwal, K.L.; Noyes, B.E.

Ann. N. Y. Acad. Sci. 343, 433-442, 1980

A:Title: Studies on gastrin mRNA structure using an oligonucleotide probe.

A:Reference number: I46622; MUID:80240380; PMID:6930858

A:Accession: I46622

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 56-82 <AGA>

A:Cross-references: GB:M25036; NID:g164626; PIDN:AAA31111.1; PID:g164627

R:Desmond, H.; Varro, A.; Young, J.; Gregory, H.; Nemeth, J.; Dockray, G.J.

Regul. Pept. 25, 223-233, 1989

A:Title: The constitution and properties of phosphorylated and unphosphorylated C-ter

A:Reference number: A60070; MUID:89331947; PMID:2756156

A:Accession: A60070

A>Status: preliminary

A:Molecule type: protein

A:Residues: 97-104 <DES>

C:Superfamily: gastrin

C:Keywords: amidated carboxyl end; phosphoprotein; pyroglutamic acid; sulfoprotein

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-56/Domain: amino-terminal propeptide #status predicted <PRO>

F:59-92/Product: big gastrin #status experimental <BGN>

F:76-92/Product: gastrin #status experimental <SGN>

F:59/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experi

F:76/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experi

F:87/Binding site: sulfate (Tyr) (covalent) (partial) #status experimental

F:92/Modified site: amidated carboxyl end (Phe) (amide in mature form from following

F:96/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 90.6%; Score 48; DB 1; Length 104;

Best Local Similarity 77.8%; Pred. No. 0.3;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEE 9

Db 76 QGPWLEEE 84

#### RESULT 6

GMDC

gastrin precursor [validated] - dog

N:Contains: big gastrin; gastrin

C:Species: Canis lupus familiaris (dog)

C:Date: 13-Jun-1983 #sequence\_revision 14-Jul-1994 #text\_change 20-Oct-2000

C:Accession: B61053; A61053; JS0425; A01620; B60070

R:Gantz, I.; Takeuchi, T.; Yamada, T.

Digestion 46, 99-104, 1990

A:Title: Cloning of canine gastrin cDNA's encoding variant amino acid sequences.

A:Reference number: A61053; MUID:91085716; PMID:2262079

A:Accession: B61053

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-104 <GAN>

A:Accession: A61053

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-84, 'T', 86-104 <GA2>  
A:Note: it is unclear whether the sequence difference results from polymorphism, multiple  
R:Bonato, C.; Eng, J.; Hulmes, J.D.; Miedel, M.; Pan, Y.C.E.; Yalow, R.S.  
Peptides 7, 689-693, 1986  
A:Title: Sequences of gastrins purified from a single antrum of dog and of goat.  
A:Reference number: JS0425; MUID:87016557; PMID:3763441  
A:Accession: JS0425  
A:Molecule type: protein  
A:Residues: 59-92 <BON>  
A:Experimental source: antral mucosa  
A:Note: about 10% of gastrin is sulfated  
R:Agarwal, K.L.; Kenner, G.W.; Sheppard, R.C.  
Experientia 25, 346-348, 1969  
A:Title: Structure and synthesis of canine gastrin.  
A:Reference number: A01620; MUID:69253357; PMID:5799207  
A:Accession: A01620  
A:Molecule type: protein  
A:Residues: 76-82, 'A', 84, 'E', 86-92 <AGA>  
R:Desmond, H.; Varro, A.; Young, J.; Gregory, G.J.  
Regul. Pept. 25, 223-233, 1989  
A:Title: The constitution and properties of phosphorylated and unphosphorylated C-terminal  
A:Reference number: A60070; MUID:89331947; PMID:2756156  
A:Accession: B60070  
A:Molecule type: protein  
A:Residues: 96-104 <DES>  
C:Comment: Big gastrin constitutes only about 5% of antral gastrin.  
C:Superfamily: gastrin  
C:Keywords: amidated carboxyl end; hormone; pancreas; phosphoprotein; pyroglutamic acid;  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:59-92/Product: big gastrin #status experimental <MAT>  
F:76-92/Product: gastrin #status experimental <MAL>  
F:76/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental  
F:76/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental  
F:87/Binding site: sulfate (Tyr) (covalent) (partial) #status experimental  
F:92/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gly  
F:96/Binding site: phosphate (Ser) (covalent) (partial) #status experimental

Query Match 88.6%; Score 48; DB 1; Length 104;  
Best Local Similarity 77.8%; Pred. No. 0.3;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
:|||||  
Db 76 QGPWVEEEE 84

RESULT 7  
GMSH  
gastrin - sheep  
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
C:Date: 13-Jun-1983 #sequence\_revision 13-Jun-1983 #text\_change 20-Mar-1998  
C:Accession: A01619  
R:Agarwal, K.L.; Beacham, J.; Bentley, P.H.; Gregory, R.A.; Kenner, G.W.; Sheppard, R.C.  
Nature 219, 614-615, 1968  
A:Title: Isolation, structure and synthesis of ovine and bovine gastrins.  
A:Reference number: A01619; MUID:68357500; PMID:5665711  
A:Accession: A01619  
A:Molecule type: protein  
A:Residues: 1-17 <AGA>  
C:Superfamily: gastrin  
C:Keywords: amidated carboxyl end; hormone; pancreas; phosphoprotein; pyroglutamic acid;  
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
F:12/Binding site: sulfate (Tyr) (covalent) (partial) #status experimental  
F:17/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 88.78%; Score 47; DB 1; Length 17;  
Best Local Similarity 77.8%; Pred. No. 0.057;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
:|||||  
Db 1 QGPWVEEEE 9

RESULT 8  
JS0426  
big gastrin - goat  
C:Contains: gastrin  
C:Species: Capra aegagrus hircus (domestic goat)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 20-Mar-1998  
C:Accession: JS0426  
R:Bonato, C.; Eng, J.; Hulmes, J.D.; Miedel, M.; Pan, Y.C.E.; Yalow, R.S.  
Peptides 7, 689-693, 1986  
A:Title: Sequences of gastrins purified from a single antrum of dog and of goat.  
A:Reference number: JS0425; MUID:87016557; PMID:3763441  
A:Accession: JS0426  
A:Molecule type: protein  
A:Residues: 1-34 <BON>  
A:Experimental source: antrum  
A:Note: about 90% of gastrin is sulfated  
C:Superfamily: gastrin  
C:Keywords: amidated carboxyl end; hormone; pancreas; pyroglutamic acid; secretagogue  
F:1-34/Product: big gastrin #status experimental <BGS>  
F:18-34/Product: gastrin #status experimental <GSN>  
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
F:18/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experi  
F:29/Binding site: sulfate (Tyr) (covalent) (partial) #status experimental  
F:34/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 88.7%; Score 47; DB 2; Length 34;  
Best Local Similarity 77.8%; Pred. No. 0.12;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
:|||||  
Db 18 QGPWVEEEE 26

RESULT 9  
GMSH  
gastrin precursor [validated] - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 31-Dec-1991 #sequence\_revision 23-Mar-1995 #text\_change 20-Oct-2000  
C:Accession: S14400; A14409; B01619; A01619  
R:Kim, S.J.; Dhm, K.N.; Kang, Y.K.; Yoo, O.J.  
DNA Seq. 1, 181-187, 1991  
A:Title: Bovine and feline gastrin cDNA sequences and the amino acid and nucleotide s  
A:Reference number: S14400; MUID:92127058; PMID:1773057  
A:Accession: S14400  
A:Molecule type: mRNA  
A:Residues: 1-104 <KIM>  
A:Cross-references: EMBL:X16581; NID:g648; PIDN:CAA34598.1; PID:g649  
R:Lund, T.; Olsen, J.; Rehfeld, J.F.  
Mol. Endocrinol. 3, 1585-1588, 1989  
A:Title: Cloning and sequencing of the bovine gastrin gene.  
A:Reference number: A41409; MUID:90114160; PMID:2608050  
A:Accession: A41409  
A:Molecule type: DNA  
A:Residues: 1-31, 'L', 33-36, 'R', 38-47, 'T', 49-73, 'N', 75-80, 'G', 82-95, 'M', 97-98, 'G', 100-  
A:Cross-references: GB:M31657; NID:g163079; PIDN:AAA30537.1; PID:g163080  
A:Note: the authors translated the codon CTG for residue 32 as Ala, AAT for residue 3  
R:Agarwal, K.L.; Beacham, J.; Bentley, P.H.; Gregory, R.A.; Kenner, G.W.; Sheppard, R  
Nature 219, 614-615, 1968  
A:Title: Isolation, structure and synthesis of ovine and bovine gastrins.  
A:Reference number: A01619; MUID:68357500; PMID:5665711  
A:Accession: B01619  
A:Molecule type: protein  
A:Residues: 76-92 <AGA>  
C:Genetics:  
A:Insertions: 71/1  
C:Superfamily: gastrin  
C:Keywords: amidated carboxyl end; hormone; pancreas; phosphoprotein; pyroglutamic ac  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:59-92/Product: big gastrin #status predicted <BGN>  
F:76-92/Product: gastrin #status experimental <GSN>  
F:59/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predic

F:76/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental  
F:87/Binding site: sulfate (Tyr) (covalent) #status experimental  
F:92/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gly

Query Match 88.7%; Score 47; DB 1; Length 104;  
Best Local Similarity 77.8%; Pred. No. 0.44;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEE 9  
Db 76 QGPWVEEE 84

## RESULT 10

T46166

MYB27 protein - Arabidopsis thaliana

N:Alternate names: protein T4D2.130

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Mar-2000

C:Accession: T46166

R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;

submitted to the Protein Sequence Database, December 1999

A:Reference number: Z23025

A:Accession: T46166

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-238 <N>A>

A:CROSS-references: EMBL:AL13295

A:Experimental source: cultivar Columbia; BAC clone T4D2

C:Genetics:

A:Map position: 3

A:Introns: 42/1; 85/2; 119/1

A>Note: T4D2.130

C:Superfamily: Arabidopsis myb-related protein 2; myb DNA-binding repeat homology

Query Match 84.9%; Score 45; DB 2; Length 238;

Best Local Similarity 87.5%; Pred. No. 2.4;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPWLEEE 9

Db 12 GPWLEED 19

## RESULT 11

A29541

little gastrin - Chinchilla brevicaudata

C:Species: Chinchilla brevicaudata, Chinchilla lanigera brevicaudata

C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 12-Apr-1995

C:Accession: A29541

R:Shinomura, Y.; Eng, J.; Yalow, R.S.

Biochem. Biophys. Res. Commun. 143, 7-14, 1987

A:Title: Chinchilla "big" and "little" gastrins.

A:Reference number: A90130; MUID:87156784; PMID:3827930

A:Accession: A29541

A:Molecule type: protein

A:Residues: 1-16 <SHI>

C:Superfamily: gastrin

Query Match 81.1%; Score 43; DB 2; Length 16;

Best Local Similarity 87.5%; Pred. No. 0.25;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEE 8

Db 1 EGPWAEEE 8

## RESULT 12

C75119

deoxyhypusine synthase (EC 2.5.1.46) dysl PAB0511 [similarity] - Pyrococcus abyssi (stra

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Nov-2001

C:Accession: C75119  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s  
A:Reference number: A75001  
A:Accession: C75119  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-335 <KAW>  
A:CROSS-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49668.1; PID:g545  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: dysl; PAB0511  
C:Superfamily: yeast deoxyhypusine synthase  
C:Keywords: hypusine biosynthesis; oxidoreductase; transferase

Query Match 81.1%; Score 43; DB 2; Length 335;  
Best Local Similarity 77.8%; Pred. No. 7.8;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEE 9

Db 20 EGPWLEDE 28

## RESULT 13

F64401

deoxyhypusine synthase (EC 2.5.1.46) MJ0814 [similarity] - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Nov-2001

C:Accession: F64401

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak

; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,

rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc

A:Reference number: A64300; MUID:96337999; PMID:8688087

A:Accession: F64401

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-370 <BUL>

A:CROSS-references: GB:U67525; GB:L77117; NID:g2826325; PIDN:AAB98813.1; PID:gl591503

C:Genetics:

A:Map position: REV738679-737567

C:Superfamily: yeast deoxyhypusine synthase

C:Keywords: hypusine biosynthesis; NAD; oxidoreductase; transferase

Query Match 79.2%; Score 42; DB 2; Length 370;

Best Local Similarity 75.0%; Pred. No. 13;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEE 8

Db 54 EGPWLEDD 61

## RESULT 14

C91080

hypothetical protein ECS3611 [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: C91080

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: C91080

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-248 <HAY>

A:CROSS-references: GB:BA000007; PIDN:BA837034.1; PID:gl3363082; GSPDB:GN00154

A: Experimental source: strain O157:H7, substrain RMD 0509952  
C: Genetics:  
A: Gene: ECs3611

Query Match 77.4% Score 41; DB 2; Length 248;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9

:||||:|

Db 234 QGPWLSKEE 242

#### RESULT 15

D85925  
hypothetical protein Z4066 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C: Species: Escherichia coli  
C: Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C: Accession: D85925  
R: Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,  
Nature 409, 529-533, 2001

A: Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A: Reference number: A85480; MUID: 21074935; PMID: 11206551

A: Accession: D85925

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-248 <STO>

A: Cross-references: GB:AE005174; NID: g12517214; PIDN: AAG57864.1; GSPDB: GN00145; UWGP: Z40

A: Experimental source: strain O157:H7, substrain EDL933

C: Genetics:

A: Gene: Z4066

Query Match 77.4% Score 41; DB 2; Length 248;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9

:||||:|

Db 234 QGPWLSKEE 242

Search completed: February 5, 2003, 09:49:47

Job time : 12.5714 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:41:15 ; Search time 6.42857 Seconds  
(without alignments)  
58.067 Million cell updates/sec

Title: US-09-700-329-1  
Perfect score: 53  
Sequence: 1 EGPWLEEEE 9 ,

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	50	94.3	33	1 GAST_DIDMA	P33713 didelphis m
2	50	94.3	101	1 GAST_HUMAN	P01350 homo sapien
3	50	94.3	104	1 GAST_FELCA	P01354 felis silve
4	48	90.6	17	1 GAST_MACMU	P33714 macaca mula
5	48	90.6	104	1 GAST_CANFA	P01353 canis fami
6	48	90.6	104	1 GAST_PIG	P01351 sus scrofa
7	47	88.7	34	1 GAST_CAPIH	P04564 capra hircu
8	47	88.7	104	1 GAST_BOVIN	P01352 bos taurus
9	47	88.7	104	1 GAST_SHEEP	O02686 ovis aries
10	46	86.8	107	1 GAST_HORSE	P55885 equus cabal
11	43	81.1	335	1 DHVS_PVRAB	Q9Y0N5 pyrococcus
12	42	79.2	351	1 DHVS_METJA	Q58224 methanococc
13	42	79.2	472	1 ZIM3_HUMAN	P09pe6 homo sapien
14	40	75.5	33	1 GAST_CAVPO	P06885 cavia porce
15	40	75.5	33	1 GAST_CHIBR	P10034 chinchilla
16	40	75.5	110	1 2SSI_BRANA	P24565 brassica na
17	39	73.6	525	1 LAG3_HUMAN	P18627 homo sapien
18	39	73.6	930	1 AT55_HUMAN	Q9una0 homo sapien
19	37	69.8	222	1 Y686_ARCFU	O29571 archaeoglob
20	37	69.8	301	1 SC14_KLULA	P24859 kluyveromyc
21	37	69.8	316	1 ERA_CAUCR	P58071 caulobacter
22	37	69.8	365	1 IA02_GORGO	P30376 gorilla gor
23	37	69.8	365	1 IA03_GORGO	P30377 gorilla gor
24	37	69.8	365	1 IA04_GORGO	P30378 gorilla gor
25	37	69.8	365	1 IA80_HUMAN	Q09160 homo sapien
26	37	69.8	382	1 MYB_AVIMB	P57016 mus musculu
27	37	69.8	528	1 LAD1_MOUSE	P01104 avian myelo
28	37	69.8	624	1 MYB_XENLA	Q08759 xenopus lae
29	37	69.8	636	1 MYB_MOUSE	P06876 mus musculu
30	37	69.8	640	1 MYB_BOVIN	P48200 bos taurus
31	37	69.8	640	1 MYB_HUMAN	P10242 homo sapien
32	37	69.8	641	1 MYB_CHICK	P10103 gallus gall
33	37	69.8	686	1 MYBB_CHICK	Q03237 gallus gall

ALIGNMENTS

RESULT 1

ID	GAST_DIDMA	STANDARD;	PRT;	33 AA.
AC	P33713:			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Big gastrin.			
GN	GAS.			
OS	Didelphis marsupialis virginiana (North American opossum).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Metatheria; Didelphimorphia; Didelphidae; Didelphis.			
OX	NCBI_TaxID=9267;			
RN	[1]			
RP	SEQUENCE.			
RX	MEDLINE=90298616; PubMed=2361360;			
RA	Shinomura Y., Eng J., Rattan S.C., Yalow R.S.;			
RA	"Opossum (Didelphis virginiana) 'little' and 'big' gastrins.";			
RL	Comp. Biochem. Physiol. 96B:239-242(1990).			
CC	FUNCTION: Gastrin stimulates the stomach mucosa to produce and secrete hydrochloric acid and the pancreas to secrete its digestive enzymes. It also stimulates smooth muscle contraction and increases blood circulation and water secretion in the stomach and intestine.			
CC	-!- SUBCELLULAR LOCATION: Secreted.			
CC	-!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.			
CC	PIR: A60506; A60506.			
DR	InterPro: IPR001651; Gastrin.			
DR	Pfam: PF00918; Gastrin; 1.			
DR	PROSITE: PS00259; GASTRIN; 1.			
KW	Hormone; Cleavage on pair of basic residues; Amidation; Sulfation.			
FT	PEPTIDE 1 33			
FT	PEPTIDE 18 33			
FT	MOD_RES 1 1			
FT	MOD_RES 18 18			
FT	MOD_RES 28 28			
FT	MOD_RES 33 33			
SQ	SEQUENCE 33 AA; 3856 MW; 217D28C15027B661 CRC64;			

Query Match 94.3%; Score 50; DB 1; Length 33;  
Best Local Similarity 88.9%; Pred. No. 0.021; Indels 0; Gaps 0;  
Matches 8; Conservative 1; Mismatches 0;

Qy 1 EGPWLEEEE 9  
:|||||||  
Db 18 QGPWLEEEE 26

RESULT 2

ID	GAST_HUMAN	STANDARD;	PRT;	101 AA.
AC	P01350; P78464; P78463;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	23-OCT-1986 (Rel. 02, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Gastrin precursor.			

GN GAS.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RX SEQUENCE FROM N.A.  
RX MEDLINE=87219893; PubMed=3034736;  
RA Kariya Y., Kato K., Hayashizaki Y., Himeno S., Tarui S., Matsubara K.;  
RT "Expression of human gastrin gene in normal and gastrinoma tissues.";  
RL Gene 50:345-352(1986).  
RN [2]  
RX SEQUENCE FROM N.A.  
RX MEDLINE=84272693; PubMed=6087340;  
RA Ito R., Sato K., Helmer T., Jay G., Agarwal K.L.;  
RT "Structural analysis of the gene encoding human gastrin: the large  
RT intron contains an Alu sequence.";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:4662-4666(1984).  
RN [3]  
RX SEQUENCE FROM N.A.  
RX MEDLINE=84169471; PubMed=6324077;  
RA Kato K., Hayashizaki Y., Takahashi Y., Himeno S., Matsubara K.;  
RT "Molecular cloning of the human gastrin gene.";  
RL Nucleic Acids Res. 11:8197-8203(1983).  
RN [4]  
RX SEQUENCE FROM N.A.  
RX MEDLINE=83221503; PubMed=6574456;  
RA Boel E., Vuust J., Norris F., Norris K., Wind A., Rehfeld J.F.,  
RA Mareker K.A.;  
RT "Molecular cloning of human gastrin cDNA: evidence for evolution of  
RT gastrin by gene duplication.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:2866-2869(1983).  
RN [5]  
RX SEQUENCE FROM N.A.  
RX MEDLINE=84144842; PubMed=6322186;  
RA Wiborg O., Berglund L., Boel E., Norris F., Norris K., Rehfeld J.F.,  
RA Mareker K.A., Vuust J.;  
RT "Structure of a human gastrin gene.";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:1067-1069(1984).  
RN [6]  
RX SEQUENCE FROM N.A.  
RX MEDLINE=84159488; PubMed=6689486;  
RA Kato K., Himeno S., Takahashi Y., Wakabayashi T., Tarui S.,  
RA Matsubara K.;  
RT "Molecular cloning of human gastrin precursor cDNA.";  
RL Gene 26:53-57(1983).  
RN [7]  
RX SEQUENCE OF 22-101, AND CHARACTERIZATION OF GASTRIN 71.  
RC TISSUE-Antral mucosa;  
RX MEDLINE=94333379; PubMed=8055952;  
RA Rehfeld J.F., Johnsen A.H.;  
RT "Identification of gastrin component I as gastrin-71. The largest  
RT possible bioactive progastrin product.";  
RL Eur. J. Biochem. 223:765-773(1994).  
RN [8]  
RX SEQUENCE OF 76-92.  
RX MEDLINE=67021327; PubMed=5921183;  
RA Bentley P.H., Kenner G.W., Sheppard R.C.;  
RT "Structures of human gastrins I and II.";  
RL Nature 209:583-585(1966).  
RN [9]  
RX SEQUENCE OF 59-68.  
RX MEDLINE=89273602; PubMed=2730647;  
RA Higashimoto Y., Himeno S., Shinomura Y., Nagao K., Tamura T.,  
RA Tarui S.;  
RT "Purification and structural determination of urinary NH2-terminal  
RT big gastrin fragments.";  
RL Biochem. Biophys. Res. Commun. 160:1364-1370(1989).  
RN [10]  
RX SEQUENCE OF 76-92.  
RX MEDLINE=69298172; PubMed=5822140;  
RA Gregory R.A., Tracy H.J., Agarwal K.L., Grossman M.I.;  
RT "Amino acid constitution of two gastrins isolated from

RT Zollinger-Ellison tumour tissue.";  
RL Gut 10:603-608(1969).  
CC -!- FUNCTION: Gastrin stimulates the stomach mucosa to produce and  
CC secrete hydrochloric acid and the pancreas to secrete its  
CC digestive enzymes. It also stimulates smooth muscle contraction  
CC and increases blood circulation and water secretion in the stomach  
CC and intestine.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- PM: TWO DIFFERENT PROCESSING PATHWAYS PROBABLY EXIST IN ANTRAL  
CC G-CELLS. IN THE DOMINANT PATHWAY PROGASTRIN IS CLEAVED AT THREE  
CC SITES RESULTING IN TWO MAJOR BIOACTIVE GASTRINS, GASTRIN-34 AND  
CC GASTRIN-17. IN THE PUTATIVE ALTERNATIVE PATHWAY, PROGASTRIN MAY  
CC BE PROCESSED ONLY AT THE MOST C-TERMINAL DIBASIC SITE RESULTING  
CC IN THE SYNTHESIS OF GASTRIN-71.  
CC -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X00183; CAA25005.1; -  
CC EMBL; X00183; CAA25006.1; -  
CC EMBL; X00183; CAA25007.1; -  
CC EMBL; V00511; CAA23769.1; -  
CC EMBL; M15958; AAA52520.1; -  
CC EMBL; K01254; AAB59533.1; -  
CC PIR; A01617; GMHUB.  
CC PIR; A18854; A18854.  
CC PIR; A32487; A32487.  
CC PIR; B32487; B32487.  
CC PIR; C32487; C32487.  
CC Gene; HGNC:4164; GAS.  
CC MIN; 137250; -  
CC InterPro; IPR001651; Gastrin.  
CC Pfam; PF00918; Gastrin; 1.  
CC SMART; SM00029; GASTRIN; 1.  
CC PROSITE; PS00259; GASTRIN; 1.  
CC Hormone; Cleavage on pair of basic residues; Amidation; Sulfation;  
CC Signal; Phosphorylation.  
CC SIGNAL 1 21  
CC FT PEPTIDE 22 92 GASTRIN 71 (COMPONENT I).  
CC FT PEPTIDE 40 92 GASTRIN 52 (POTENTIAL).  
CC FT PEPTIDE 59 92 BIG GASTRIN (GASTRIN 34) (COMPONENT II).  
CC FT PEPTIDE 76 92 GASTRIN (GASTRIN 17) (COMPONENT III).  
CC FT MOD\_RES 59 59 PYRROLIDONE CARBOXYLIC ACID.  
CC FT MOD\_RES 76 76 PYRROLIDONE CARBOXYLIC ACID.  
CC FT MOD\_RES 87 87 SULFATION (IN 33% OF THE CHAINS).  
CC FT MOD\_RES 92 92 AMIDATION (G-93 PROVIDE AMIDE GROUP).  
CC FT MOD\_RES 96 96 PHOSPHORYLATION (BY SIMILARITY).  
CC SQ SEQUENCE 101 AA; 11394 MW; A03C847FCFE7216C CRC64;  
  
CC Query Match 94.3%; Score 50; DB 1; Length 101;  
CC Best Local Similarity 88.9%; Pred. No. 0.068;  
CC Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
CC Qy 1 EGPWLEEEE 9  
CC Db 76 QGPWLEEEE 84  
  
CC RESULT 3  
CC GAST\_FELCA  
CC ID GAST\_FELCA STANDARD; PRT; 104 AA.  
CC AC P01354;  
CC DT 21-JUL-1986 (Rel. 01, Created)  
CC DT 01-NOV-1991 (Rel. 20, Last sequence update)  
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
CC DE Gastrin precursor.  
CC GN GAS.

```
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92127058; PubMed=1773057;
RA Kim S.J., Uhm K.N., Kang Y.K., Yoo O.J.;
RT "Bovine and feline gastrin cDNA sequences and the amino acid and
RT nucleotide sequence homologies among mammalian species.";
RL DNA Seq. 1:181-187(1991).
RN [2]
RP SEQUENCE OF 76-92.
RX MEDLINE=69206035; PubMed=5784957;
RA Agarwal K.L., Kenner G.W., Sheppard R.C.;
RT "Feline gastrin. An example of peptide sequence analysis by mass
RT spectrometry.";
RL J. Am. Chem. Soc. 91:3096-3097(1969).
CC -!- FUNCTION: Gastrin stimulates the stomach mucosa to produce and
CC secrete hydrochloric acid and the pancreas to secrete its
CC digestive enzymes. It also stimulates smooth muscle contraction
CC and increases blood circulation and water secretion in the stomach
CC and intestine.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16582; CAA34599.1; -
DR PIR: A01621; GMCT.
DR PIR: S14401; S14401.
DR InterPro: IPR001651; Gastrin.
DR Pfam: PF00918; Gastrin; 1.
DR SMART: SM00029; GASTRIN; 1.
DR PROSITE: PS00259; GASTRIN; 1.
KW Hormone; Cleavage on pair of basic residues; Amidation; Sulfation;
KW Signal; Phosphorylation.
FT SIGNAL 1 21
FT PEPTIDE 59 92 BIG GASTRIN (GASTRIN 34).
FT PEPTIDE 76 92 GASTRIN.
FT MOD_RES 59 59 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 76 76 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 87 87 SULFATION (PARTIAL).
FT MOD_RES 92 92 AMIDATION (G-93 PROVIDE AMIDE GROUP).
FT MOD_RES 96 96 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 104 AA; 11482 MW; 4DB92E4416A7AC9F CRC64;

Query Match 94.3%; Score 50; DB 1; Length 104;
Best Local Similarity 88.9%; Pred. No. 0.07;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEE 9
Db 76 QGPWLEEE 84
:|||||||
:|||||||

RESULT 4
GAST_MACMU STANDARD; PRT; 17 AA.
AC F33714;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Gastrin.
GN GAS.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE.
RX MEDLINE=91164506; PubMed=2003150;
RA Yu J., Xin Y., Eng J., Yalow R.S.;
RT "Rhesus monkey gastroenteropancreatic hormones: relationship to human
RT sequences.";
RL Regul. Pept. 32:39-45(1991).
CC -!- FUNCTION: Gastrin stimulates the stomach mucosa to produce and
CC secrete hydrochloric acid and the pancreas to secrete its
CC digestive enzymes. It also stimulates smooth muscle contraction
CC and increases blood circulation and water secretion in the stomach
CC and intestine.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR PIR: A60071; A60071.
DR InterPro: IPR001651; Gastrin.
DR PROSITE: PS00259; GASTRIN; 1.
KW Hormone; Amidation; Sulfation.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 12 12 SULFATION.
FT MOD_RES 17 17 AMIDATION.
SQ SEQUENCE 17 AA; 2076 MW; 6F6E92C73611D39A CRC64;

Query Match 90.6%; Score 48; DB 1; Length 17;
Best Local Similarity 77.8%; Pred. No. 0.024;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEE 9
Db 1 QGPWLEEE 9
:|||||
:|||||

RESULT 5
GAST_CANFA STANDARD; PRT; 104 AA.
AC P01353;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Gastrin precursor.
GN GAS.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-Antral mucosa;
RX MEDLINE=91085716; PubMed=2262079;
RA Gantz I., Takeuchi T., Yamada T.;
RT "Cloning of canine gastrin cDNA's encoding variant amino acid
RT sequences.";
RL Digestion 46:99-104(1990).
RN [2]
RP SEQUENCE OF 59-92.
RX TISSUE-Antral mucosa;
RX MEDLINE=87016557; PubMed=3763441;
RA Bonato C., Eng J., Hulmes J.D., Miesel M., Pan Y.-C.E., Yalow R.S.;
RT "Sequences of gastrins purified from a single antrum of dog and of
RT goat.";
RL Peptides 7:689-693(1986).
RN [3]
RP SEQUENCE OF 76-92, AND PHOSPHORYLATION.
RX MEDLINE=69253357; PubMed=5799207;
RA Agarwal K.L., Kenner G.W., Sheppard R.C.;
RT "Structure and synthesis of canine gastrin.";
RL Experientia 25:346-348(1969).
RN [4]
RP SEQUENCE OF 96-104.
RX TISSUE-Antral mucosa;
RC
```

```
RX MEDLINE=89331947; PubMed=2756156;
RA Desmond H., Varro A., Young J., Gregory H., Nemeth J., Dockray G.J.;
RT "The constitution and properties of phosphorylated and
RT unphosphorylated C-terminal fragments of progastrin from dog and
RT ferret antrum";
RL Regul. Pept. 25:223-233(1989).
CC -!- FUNCTION: Gastrin stimulates the stomach mucosa to produce and
CC secrete hydrochloric acid and the pancreas to secrete its
CC digestive enzymes. It also stimulates smooth muscle contraction
CC and increases blood circulation and water secretion in the stomach
CC and intestine.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR PIR; JS0425; GMDG.
DR InterPro; IPR001651; Gastrin.
DR Pfam; PF00918; Gastrin; 1.
DR SMART; SM00029; GASTRIN; 1.
DR PROSITE; PS00259; GASTRIN; 1.
KW Hormone; Cleavage on pair of basic residues; Amidation; Sulfation;
KW Signal; Phosphorylation.
FT SIGNAL 1 21
FT PEPTIDE 59 92 POTENTIAL.
FT PEPTIDE 76 92 BIG GASTRIN (GASTRIN 34).
FT MOD_RES 59 59 GASTRIN.
FT MOD_RES 76 76 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 87 87 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 92 92 SULFATION (PARTIAL).
FT MOD_RES 96 96 AMIDATION (G-93 PROVIDE AMIDE GROUP).
FT VARIANT 85 85 A -> T.
FT CONFLICT 83 85 EEA -> AEE (IN REF. 3).
SQ SEQUENCE 104 AA; 11519 MW; 73BF72A18DFE78CA CRC64;

Query Match 90.6%; Score 48; DB 1; Length 104;
Best Local Similarity 77.8%; Pred. No. 0.15;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EGPWLEEEE 9
DB 76 QGPWMEEEE 84

RESULT 6
GAST_PIG STANDARD; PRT; 104 AA.
AC P01351;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Gastrin precursor.
GN GAS.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82174533; PubMed=69511161;
RA Yoo O.J., Powell C.T., Agarwal K.L.;
RT "Molecular cloning and nucleotide sequence of full-length of cDNA
RT coding for porcine gastrin.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:1049-1053(1982).
RN [2]
RP SEQUENCE OF 56-82 FROM N.A.
RX MEDLINE=80240380; PubMed=6930858;
RA Agarwal K.L., Noyes B.E.;
RT "Studies on gastrin mRNA structure using an oligonucleotide probe.";
RL Ann. N.Y. Acad. Sci. 343:433-442(1980).
RN [3]
RP SEQUENCE OF 76-92.
RA Gregory H., Hardy P.M., Jones D.S., Kenner G.W., Sheppard R.C.;
RT "The antral hormone gastrin.";
RL Nature 204:931-933(1964).
RN [4]
```

```
RP SYNTHESIS.
RA Anderson J.C., Barton M.A., Gregory R.A., Hardy P.M., Kenner G.W.,
RA McLeod J.K., Preston J., Sheppard R.C., Morley J.S.;
RT "Synthesis of gastrin.";
RL Nature 204:933-934(1964).
CC -!- FUNCTION: Gastrin stimulates the stomach mucosa to produce and
CC secrete hydrochloric acid and the pancreas to secrete its
CC digestive enzymes. It also stimulates smooth muscle contraction
CC and increases blood circulation and water secretion in the stomach
CC and intestine.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR EMBL; V01303; CAA24610.1; -
DR EMBL; M25036; AAA31111.1; -
DR PIR; A01618; GMPGB.
DR InterPro; IPR001651; Gastrin.
DR Pfam; PF00918; Gastrin; 1.
DR SMART; SM00029; GASTRIN; 1.
DR PROSITE; PS00259; GASTRIN; 1.
KW Hormone; Cleavage on pair of basic residues; Amidation; Sulfation;
KW Signal; Phosphorylation.
FT SIGNAL 1 21
FT PEPTIDE 59 92 PROBABLE.
FT PEPTIDE 76 92 BIG GASTRIN (GASTRIN 34).
FT MOD_RES 59 59 GASTRIN.
FT MOD_RES 76 76 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 87 87 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 92 92 SULFATION (IN 66% OF THE CHAINS).
FT MOD_RES 96 96 AMIDATION (G-93 PROVIDE AMIDE GROUP).
FT MOD_RES 96 96 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 104 AA; 11558 MW; B0BD1D7E05304B79 CRC64;

Query Match 90.6%; Score 48; DB 1; Length 104;
Best Local Similarity 77.8%; Pred. No. 0.15;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EGPWLEEEE 9
DB 76 QGPWMEEEE 84

RESULT 7
GAST_CAPI STANDARD; PRT; 34 AA.
AC P04564;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Big gastrin.
GN GAS.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE.
RC TISSUE=Antral mucosa;
RX MEDLINE=87016557; PubMed=3763441;
RA Bonato C., Eng J., Hulmes J.D., Miedel M., Pan Y.-C.E., Yalow R.S.;
RT "Sequences of gastrins purified from a single antrum of dog and of
RT goat.";
RL Peptides 7:689-693(1986).
CC -!- FUNCTION: Gastrin stimulates the stomach mucosa to produce and
CC secrete hydrochloric acid and the pancreas to secrete its
```

digestive enzymes. It also stimulates smooth muscle contraction and increases blood circulation and water secretion in the stomach and intestine.

-!- SUBCELLULAR LOCATION: Secreted.

-!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.

PIR; JS0426; JS0426.

InterPro; IPR001651; Gastrin.

Pfam; PF00918; Gastrin; 1.

Hormone; Cleavage on pair of basic residues; Amidation; Sulfation.

PEPTIDE 1 34 BIG GASTRIN (GASTRIN 34).

MOD\_RES 18 34 PYRROLIDONE CARBOXYLIC ACID.

MOD\_RES 18 18 PYRROLIDONE CARBOXYLIC ACID.

MOD\_RES 29 29 SULFATION.

MOD\_RES 34 34 AMIDATION.

SEQUENCE 34 AA; 3903 MW; 67501111E76D0CF4 CRC64;

Query Match 88.7%; Score 47; DB 1; Length 34;  
Best Local Similarity 77.8%; Pred. No. 0.071;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9

Db 18 QGPWVEEEE 26

#### RESULT 8

##### GAST\_BOVIN

ID GAST\_BOVIN STANDARD; PRT; 104 AA.

AC P01352; Q28114;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Gastrin precursor.

GN GAS.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;

OC Bovidae; Bovinae; Bos.

OX NCBI\_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=90114160; PubMed=2608050;

RA Lund T., Olsen J., Rehfeld J.F.;

RT "Cloning and sequencing of the bovine gastrin gene.;"

RL Mol. Endocrinol. 3:1585-1588(1989).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=92127058; PubMed=1773057;

RA Kim S.J., Uhm K.N., Kang Y.K., Yoo O.J.;

RT "Bovine and feline gastrin cDNA sequences and the amino acid and

RNA nucleotide sequence homologies among mammalian species.;"

RL DNA Seq. 1:181-187(1991).

RN [3]

RP SEQUENCE OF 76-92.

RX MEDLINE=68357500; PubMed=5665711;

RA Agarwal K.L., Beacham J., Bentley P.H., Gregory R.A., Kenner G.W.,

RA Sheppard R.C., Tracy H.J.;

RT "Isolation, structure and synthesis of ovine and bovine gastrins.;"

RL Nature 219:614-615(1968).

CC -!- FUNCTION: Gastrin stimulates the stomach mucosa to produce and

CC secrete hydrochloric acid and the pancreas to secrete its

CC digestive enzymes. It also stimulates smooth muscle contraction

CC and increases blood circulation and water secretion in the stomach

CC and intestine.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M31657; AAA30537.1; -

DR EMBL; X16581; CAA34598.1; -

DR PIR; B01619; GMB0.

DR PIR; S14400; S14400.

DR InterPro; IPR001651; Gastrin.

DR Pfam; PF00918; Gastrin; 1.

DR SMART; SM00029; GASTRIN; 1.

DR PROSITE; PS00259; GASTRIN; 1.

KW Hormone; Cleavage on pair of basic residues; Amidation; Sulfation;

KW Signal; Phosphorylation.

FT SIGNAL 1 21

FT PEPTIDE 59 92

FT MOD\_RES 76 92

FT MOD\_RES 59 59

FT MOD\_RES 76 76

FT MOD\_RES 87 87

FT MOD\_RES 92 92

FT CONFLICT 32 32

FT CONFLICT 37 37

FT CONFLICT 48 48

FT CONFLICT 74 74

FT CONFLICT 81 81

FT CONFLICT 96 96

FT CONFLICT 99 99

FT CONFLICT 104 AA; 11573 MW; 54D03BF200F299F2 CRC64;

SEQ SEQUENCE 104 AA; 11573 MW; 54D03BF200F299F2 CRC64;

Query Match 88.7%; Score 47; DB 1; Length 104;

Best Local Similarity 77.8%; Pred. No. 0.22;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9

Db 76 QGPWVEEEE 84

#### RESULT 9

##### GAST\_SHEEP

ID GAST\_SHEEP STANDARD; PRT; 104 AA.

AC 002686;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Gastrin precursor.

GN GAS.

OS Ovis aries (Sheep).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;

OC Bovidae; Caprinae; Ovis.

OX NCBI\_TaxID=9940;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98182586; PubMed=9522119;

RA Moore C., Jie R., Shulkes A., Baldwin G.S.;

RT "Molecular cloning and sequence of the ovine gastrin gene.;"

RL DNA Seq. 8:39-44(1997).

RN [2]

RP SEQUENCE OF 76-92.

RX MEDLINE=68357500; PubMed=5665711;

RA Agarwal K.L., Beacham J., Bentley P.H., Gregory R.A., Kenner G.W.,

RA Sheppard R.C., Tracy H.J.;

RT "Isolation, structure and synthesis of ovine and bovine gastrins.;"

RL Nature 219:614-615(1968).

CC -!- FUNCTION: Gastrin stimulates the stomach mucosa to produce and

CC secrete hydrochloric acid and the pancreas to secrete its

CC digestive enzymes. It also stimulates smooth muscle contraction

CC and increases blood circulation and water secretion in the stomach

CC and intestine.

```

CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U92801; AAB51307.1; -.
CC PIR: A01619; GMSH.
CC InterPro: IPR001651; Gastrin.
CC Pfam: PF00918; Gastrin; 1.
CC SMART: SM00029; Gastrin; 1.
CC PROSITE: PS00259; Gastrin; 1.
CC Hormone; Cleavage on pair of basic residues; Amidation; Sulfation;
CC Signal; Phosphorylation.
CC -----
CC FT SIGNAL 1 21
CC FT PEPTIDE 59 92 BIG GASTRIN (GASTRIN 34).
CC FT PEPTIDE 76 92 GASTRIN.
CC FT MOD_RES 59 59 PYRROLIDONE CARBOXYLIC ACID.
CC FT MOD_RES 76 76 PYRROLIDONE CARBOXYLIC ACID.
CC FT MOD_RES 87 87 SULFATION (PARTIAL).
CC FT MOD_RES 92 92 AMIDATION (G-93 PROVIDE AMIDE GROUP).
CC FT MOD_RES 96 96 PHOSPHORYLATION (BY SIMILARITY).
CC SEQUENCE 104 AA; 11532 MW; 624063D4B5CE5AFD CRC64;

Query Match 88.7%; Score 47; DB 1; Length 104;
Best Local Similarity 77.8%; Pred. No. 0.22;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
DB 76 QGPWLEEEE 84

RESULT 10
GAST_HORSE
ID GAST_HORSE STANDARD; PRT; 107 AA.
AC P55885;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Gastrin precursor.
GN GAS.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Antral mucosa;
RA Johnson A.H., Sandlin A., Rourke I.J., Bundgaard J.R., Nilsson G.,
RA Rehfeld J.F.;
RT Unique progastrin processing in equine G-cells suggests marginal
RT tyrosyl sulfotransferase activity.
RL Eur. J. Biochem. 255:432-438(1998).
CC -!- FUNCTION: Gastrin stimulates the stomach mucosa to produce and
CC secrete hydrochloric acid and the pancreas to secrete its
CC digestive enzymes. It also stimulates smooth muscle contraction
CC and increases blood circulation and water secretion in the stomach
CC and intestine.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y09440; CAA70590.1; -.
CC InterPro: IPR001651; Gastrin.
CC Pfam: PF00918; Gastrin; 1.
CC SMART: SM00029; Gastrin; 1.
CC PROSITE: PS00259; Gastrin; 1.
CC Hormone; Cleavage on pair of basic residues; Amidation; Sulfation;
CC Signal; Phosphorylation.
CC -----
CC FT SIGNAL 1 21
CC FT PEPTIDE 62 95 BIG GASTRIN (GASTRIN 34).
CC FT PEPTIDE 79 95 GASTRIN.
CC FT MOD_RES 62 62 PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).
CC FT MOD_RES 79 79 PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).
CC FT MOD_RES 90 90 SULFATION (PARTIAL).
CC FT MOD_RES 95 95 SULFATION (G-96 PROVIDE AMIDE GROUP).
CC FT MOD_RES 99 99 PHOSPHORYLATION (BY SIMILARITY).
CC SEQUENCE 107 AA; 11884 MW; 104166CAAE5C234F CRC64;

Query Match 86.8%; Score 46; DB 1; Length 107;
Best Local Similarity 77.8%; Pred. No. 0.34;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
DB 79 QGPWLEEEE 87

RESULT 11
DHS_PYRAB
ID DHS_PYRAB STANDARD; PRT; 335 AA.
AC Q9V0N5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable deoxyhypusine synthase (EC 2.5.1.46) (DHS).
GN DYS OR PAR0511.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE NAD-DEPENDENT OXIDATIVE CLEAVAGE OF
CC SPERMIDINE AND THE SUBSEQUENT TRANSFER OF THE BUTYLAMINE MOIETY OF
CC SPERMIDINE TO THE EPSILON-AMINO GROUP OF A SPECIFIC LYSINE RESIDUE
CC OF THE EIF-5A PRECURSOR PROTEIN TO FORM THE INTERMEDIATE
CC DEOXYHYPUSINE RESIDUE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: [eif5a-precursor]-lysine + spermidine =
CC [eif5a-precursor]-deoxyhypusine + propane-1,3-diamine.
CC -!- COFACTOR: NAD (BY SIMILARITY).
CC -!- PATHWAY: Hypusine biosynthesis; first step.
CC -!- SIMILARITY: BELONGS TO THE DEOXYHYPUSINE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ248285; CAB49668.1; -.
CC HSSP: P49366; 1DHS.
CC InterPro: IPR002773; DS.
CC Pfam: PF01916; DS; 1.
CC ProDom: PD007730; DS; 1.

```

DR TIGR00321; dhys; 1.  
 KW Hypusine biosynthesis; Transferase; NAD; Complete proteome.  
 SQ SEQUENCE 335 AA; 37997 MW; 3DE853F38029EC55 CRC64;

Query Match 81.1%; Score 43; DB 1; Length 335;  
 Best Local Similarity 77.8%; Pred. No. 3.5;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEE 9  
 |||||:|  
 Db 20 EGPWLEVE 28 {

## RESULT 12

DHYS\_METJA STANDARD; PRT; 351 AA.

AC Q58224;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-NOV-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable deoxyhypusine synthase (EC 2.5.1.46) (DHS).  
 GN DYS OR MJ0814.

OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcales; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;

RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;

RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kervilange A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klek H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RL Science 273:1058-1073(1996).

CC -!- FUNCTION: CATALYZES THE NAD-DEPENDENT OXIDATIVE CLEAVAGE OF  
 CC SPERMIDINE AND THE SUBSEQUENT TRANSFER OF THE BUTYLAMINE MOIETY OF  
 CC SPERMIDINE TO THE EPSILON-AMINO GROUP OF A SPECIFIC LYSINE RESIDUE  
 CC OF THE EIF-5A PRECURSOR PROTEIN TO FORM THE INTERMEDIATE  
 CC DEOXYHYPUSINE RESIDUE (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: [eif5A-precursor]-lysine + spermidine =  
 CC [eif5A-precursor]-deoxyhypusine + propane-1,3-diamine.  
 CC -!- COFACTOR: NAD (BY SIMILARITY).  
 CC -!- PATHWAY: Hypusine biosynthesis; first step.

CC -!- SIMILARITY: BELONGS TO THE DEOXYHYPUSINE SYNTHASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U67525; AAB98813.1; ALT\_INIT.  
 DR HSSP; P49366; LDHS.

DR TIGR; MJ0814; -

DR InterPro; IPR002773; DS.

DR InterPro; IPR001230; Prenyl\_site.

DR Pfam; PF01916; DS; 1.

DR ProDom; PD007730; DS; 1.

DR TIGRFAMs; TIGR00321; dhys; 1.

KW Hypusine biosynthesis; Transferase; NAD; Complete proteome.  
 SQ SEQUENCE 351 AA; 39978 MW; 7F7B306337FED7CD CRC64;

Query Match 79.2%; Score 42; DB 1; Length 351;  
 Best Local Similarity 75.0%; Pred. No. 5.4;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEE 8  
 |||||:|

Db 35 EGPWLEDD 42

## RESULT 13

ZIM3\_HUMAN

ID ZIM3\_HUMAN STANDARD; PRT; 472 AA.

AC Q96PE6; ZIM3

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Zinc finger imprinted 3.

GN ZIM3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21429118; PubMed=11543637;

RA Kim J., Bergmann A., Wehrli E., Lu X., Stubbs L.;

RT "Imprinting and evolution of two Kruppel-type zinc-finger genes, ZIM3  
 and ZNF264, located in the PEG3/USP29 imprinted domain.";  
 RL Genomics 77:91-98(2001).

CC -!- FUNCTION: May function as a transcription factor.

CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-

CC FINGER PROTEINS.

CC -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF365931; AAL1635.1;

CC Genew; HGNC:16366; ZIM3.

CC InterPro; IPR001909; KRAB.

CC InterPro; IPR000822; Znf\_C2H2.

CC Pfam; PF00096; zf-C2H2; 1.

CC Pfam; PF01352; KRAB; 1.

CC PROSITE; PS00805; KRAB; 1.

CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 11.

CC PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 11.

CC KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;

CC DNA-binding; Repeat.

CC DOMAIN 80

CC DOMAIN 167 470

CC ZINC\_FINGERS.

CC ZN\_FING 167 189

CC C2H2-TYPE.

CC ZN\_FING 195 217

CC C2H2-TYPE.

CC ZN\_FING 223 245

CC C2H2-TYPE.

CC ZN\_FING 251 273

CC C2H2-TYPE.

CC ZN\_FING 279 301

CC C2H2-TYPE.

CC ZN\_FING 307 329

CC C2H2-TYPE.

CC ZN\_FING 335 357

CC C2H2-TYPE.

CC ZN\_FING 363 385

CC C2H2-TYPE.

CC ZN\_FING 391 413

CC C2H2-TYPE.

CC ZN\_FING 419 441

CC C2H2-TYPE.

CC ZN\_FING 447 470

CC C2H2-TYPE.

SQ SEQUENCE 472 AA; 54498 MW; 755279B78653F286 CRC64;

Query Match

Best Local Similarity 79.2%; Score 42; DB 1; Length 472;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PWLEEEE 9

|||||

```

Db 67 PWLEEE 73
RESULT 14
GAST_CAVPO STANDARD; PRT; 33 AA.
AC P06885;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Big gastrin.
GN GAS.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE.
RX MEDLINE=86309993; PubMed=3747718;
RA Bonato C., Eng J., Pan Y.-C.E., Miedel M., Hulmes J.D., Yalow R.S.;
RT "Guinea pig 33-amino acid gastrin.";
RL Life Sci. 39:959-964(1986).
CC -!- FUNCTION: Gastrin stimulates the stomach mucosa to produce and
CC secrete hydrochloric acid and the pancreas to secrete its
CC digestive enzymes. It also stimulates smooth muscle contraction
CC and increases blood circulation and water secretion in the stomach
CC and intestine.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR PIR; A26089; GMPB.
DR InterPro: IPR001651; Gastrin.
DR Pfam: PF00918; Gastrin; 1.
DR PROSITE: PS00259; GASTRIN; 1.
KW Hormone; Cleavage on pair of basic residues; Amidation
FT PEPTIDE 1 33 BIG GASTRIN (GASTRIN 33)..
FT PEPTIDE 18 33 GASTRIN.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 18 18 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 33 33 AMIDATION.
SQ SEQUENCE 33 AA; 3757 MW; B37C251CD40EB30C CRC64;

Query Match 75.5%; Score 40; DB 1; Length 33;
Best Local Similarity 75.0%; Pred. No. 1;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EGPWLEEE 8
Db 18 QGPWAEED 25

RESULT 15
GAST_CHIBR STANDARD; PRT; 33 AA.
AC P10034;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Big gastrin.
GN GAS.
OS Chinchilla brevicaudata (Chinchilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Chinchillidae;
OC Chinchilla.
OX NCBI_TaxID=10152;
RN [1]
RP SEQUENCE.
RX MEDLINE=87156784; PubMed=3827930;
RA Shinomura Y., Eng J., Yalow R.S.;
RT "Chinchilla 'big' and 'little' gastrins.";
RL Biochem. Biophys. Res. Commun. 143:7-14(1987).
CC -!- FUNCTION: Gastrin stimulates the stomach mucosa to produce and
CC secrete hydrochloric acid and the pancreas to secrete its
CC digestive enzymes. It also stimulates smooth muscle contraction

```

```

CC and increases blood circulation and water secretion in the stomach
CC and intestine.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE GASTRIN/CHOLECYSTOKININ FAMILY.
DR PIR; A29541; A29541.
DR PIR; B29541; B29541.
DR InterPro: IPR001651; Gastrin.
DR Pfam: PF00918; Gastrin; 1.
DR PROSITE: PS00259; GASTRIN; 1.
KW Hormone; Cleavage on pair of basic residues; Amidation; Sulfation.
FT PEPTIDE 1 33 BIG GASTRIN (GASTRIN 33)..
FT PEPTIDE 18 33 GASTRIN.
FT MOD_RES 18 18 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 28 28 SULFATION.
FT MOD_RES 33 33 AMIDATION.
SQ SEQUENCE 33 AA; 3715 MW; 6F11F5CDC50FAA2D CRC64;

Query Match 75.5%; Score 40; DB 1; Length 33;
Best Local Similarity 75.0%; Pred. No. 1;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EGPWLEEE 8
Db 18 QGPWAEED 25

Search completed: February 5, 2003, 09:48:17
Job time : 7.42857 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:45:51 ; Search time 21 seconds  
(without alignments)  
88.306 Million cell updates/sec

Title: US-09-700-329-1  
Perfect score: 53  
Sequence: 1 EGPWLEEE 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	84.9	238	10 Q9SCP1	Q9SCP1 arabidopsis
2	41	77.4	160	10 Q8S415	Q8S415 zea mays (m
3	41	77.4	248	16 Q8X4R9	Q8X4R9 escherichia
4	41	77.4	273	10 Q94IF9	Q94IF9 arabidopsis
5	41	77.4	273	10 Q94IF8	Q94IF8 arabidopsis
6	41	77.4	273	10 Q94IF6	Q94IF6 arabidopsis
7	41	77.4	273	10 Q93V20	Q93V20 arabidopsis
8	41	77.4	273	10 Q93V46	Q93V46 arabidopsis
9	41	77.4	273	10 Q39028	Q39028 arabidopsis
10	41	77.4	293	10 Q49020	Q49020 gossypium h
11	41	77.4	293	10 Q94JN5	Q94JN5 gossypium h
12	41	77.4	293	10 Q94JN4	Q94JN4 gossypium h
13	40	75.5	268	10 P81394	P81394 antirrhinum
14	40	75.5	310	10 Q9FJP2	Q9FJP2 arabidopsis
15	40	75.5	311	11 Q9CQD5	Q9CQD5 mus musculu
16	40	75.5	311	11 Q8VE83	Q8VE83 mus musculu

17	40	75.5	349	10 Q9SPG8	Q9SPG8 arabidopsis
18	40	75.5	745	10 Q23618	Q23618 arabidopsis
19	40	75.5	834	12 Q85065	Q85065 peanut stn
20	39	73.6	181	17 Q8ZT18	Q8ZT18 pyrobaculum
21	39	73.6	204	7 Q31205	Q31205 mus musculu
22	39	73.6	306	10 Q8W0D0	Q8W0D0 oryza sativ
23	39	73.6	344	7 Q9BCZ1	Q9BCZ1 mus musculu
24	39	73.6	377	7 Q8WLM9	Q8WLM9 mus musculu
25	39	73.6	379	7 Q31208	Q31208 mus musculu
26	39	73.6	379	7 Q31615	Q31615 mus musculu
27	39	73.6	406	7 Q31206	Q31206 mus musculu
28	39	73.6	718	10 Q947V1	Q947V1 oryza sativ
29	39	73.6	749	5 Q8T3S1	Q8T3S1 drosophila
30	39	73.6	880	5 Q9VNU9	Q9VNU9 drosophila
31	39	73.6	960	10 Q22214	Q22214 arabidopsis
32	39	73.6	1108	10 Q8W4F7	Q8W4F7 arabidopsis
33	39	73.6	1480	4 Q96Q04	Q96Q04 homo sapien
34	38	71.7	157	10 Q8S416	Q8S416 zea mays (m
35	38	71.7	196	4 Q96M28	Q96M28 homo sapien
36	38	71.7	220	10 Q9FEV3	Q9FEV3 oryza sativ
37	38	71.7	221	10 Q94CT3	Q94CT3 oryza sativ
38	38	71.7	234	10 Q50069	Q50069 arabidopsis
39	38	71.7	235	10 Q9FN86	Q9FN86 arabidopsis
40	38	71.7	248	16 Q8ZMF0	Q8ZMF0 salmoneilla
41	38	71.7	256	10 Q9LX82	Q9LX82 arabidopsis
42	38	71.7	256	10 Q941G1	Q941G1 arabis gemm
43	38	71.7	278	10 P93391	P93391 nicotiana t
44	38	71.7	318	2 Q9X3F7	Q9X3F7 zymomonas m
45	38	71.7	325	10 Q9LFE1	Q9LFE1 arabidopsis

## ALIGNMENTS

## RESULT 1

Q9SCP1 ID Q9SCP1 PRELIMINARY: PRT; 238 AA.  
AC Q9SCP1;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE MYB27 protein.  
GN T4D2.130.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nvakatura G., Fartmann B., Dauner D., Sterr W., Holland R.,  
RA Weichselgartner M., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,  
RA Salanoubat M.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: CONTAINS 2 MYB-LIKE DOMAINS.  
DR EMBL; ALI32958; CAB64223.1; -.  
DR HSSP; P01103; IPOM.  
DR InterPro; IPR001005; Myb\_DNA\_binding.  
DR Pfam; PF00249; myb\_DNA-binding; 2.  
DR SMART; SM00395; SANT; 2.  
DR PROSITE; PS50090; MYB\_3; 2.  
KW DNA-binding; Nuclear protein.  
SQ SEQUENCE 238 AA; 27996 MW; B6DDCF70E22DE62E CRC64;

Query Match 84.9%; Score 45; DB 10; Length 238;  
Best Local Similarity 87.5%; Pred. No. 7.8;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 GPWLEEEE 9
DB 12 GPWLEED 19

RESULT 2
Q8S415 PRELIMINARY; PRT; 160 AA.
AC Q8S415:
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE P-type R2R3 Myb protein.
GN MYB49.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Jiang C., Peterson T.;
RT "Ordered Origin of the Typical Two- and Three-Repeat Myb Genes.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF470090; AAL90658.1; -.
SQ SEQUENCE 160 AA; 17920 MW; 1FCBA42C22AF1B89 CRC64;

Query Match 77.4%; Score 41; DB 10; Length 160;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
DB 15 KGPWTEED 23

RESULT 3
Q8X4R9 PRELIMINARY; PRT; 248 AA.
AC Q8X4R9:
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Orf, hypothetical protein.
GN Z4066 OR ECS3611.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouisis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RN Nature 409:529-533(2001).
RP SEQUENCE FROM N.A.
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12";
RN DNA Res. 8:111-22(2001).
DR EMBL; AE005503; AAG57864.1; -.

DR EMBL; AP002562; BAB37034.1; -.
KW Complete proteome.
SQ SEQUENCE 248 AA; 28619 MW; D0F452DA2299E3D8 CRC64;

Query Match 77.4%; Score 41; DB 16; Length 248;
Best Local Similarity 66.7%; Pred. No. 37;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
DB 234 QGPWLSKEE 242

RESULT 4
Q94IF9 PRELIMINARY; PRT; 273 AA.
AC Q94IF9:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE MYB transcription factor Atmyb2.
GN AtMYB2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Kamiya T., Kawabe A., Miyashita N.T.;
RT "Nucleotide polymorphism at the Atmyb2 locus of the wild plant
RT Arabidopsis thaliana.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB052234; BAB62114.1; -.
DR InterPro; IPR001395; Aldo/ket_red.
DR Pfam; PF00249; myb_DNA-binding; 2.
DR PROSITE; PS00063; ALDO-KETO-REDUCTASE_3; UNKNOWN_1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR PROSITE; PS00334; MYB_2; UNKNOWN_1.
DR PROSITE; PS00090; MYB_3; 2.
SQ SEQUENCE 273 AA; 31416 MW; 737F99A3C287049E CRC64;

Query Match 77.4%; Score 41; DB 10; Length 273;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
DB 22 KGPWTEED 30

RESULT 5
Q94IF8 PRELIMINARY; PRT; 273 AA.
AC Q94IF8:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE MYB transcription factor Atmyb2.
GN AtMYB2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Kamiya T., Kawabe A., Miyashita N.T.;
RT "Nucleotide polymorphism at the Atmyb2 locus of the wild plant
RT Arabidopsis thaliana.";

```

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB052235; BAB62115.1; -  
 DR InterPro; IPR001395; Aldo/ket\_red.  
 DR Pfam; PF00249; myb\_DNA-binding; 2  
 DR PROSITE; PS00063; ALDOKETO\_REDUCTASE\_3; UNKNOWN\_1.  
 DR PROSITE; PS00037; MYB\_1; UNKNOWN\_1.  
 DR PROSITE; PS00334; MYB\_2; UNKNOWN\_1.  
 DR PROSITE; PS50090; MYB\_3; 2.  
 SQ SEQUENCE 273 AA; 31429 MW; 52B5D2466790B9E6 CRC64;

Query Match 77.4%; Score 41; DB 10; Length 273;  
 Best Local Similarity 66.7%; Pred. No. 40;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
 :||| |||:  
 Db 22 KGPWTEED 30

## RESULT 6

Q94IF6 PRELIMINARY; PRT; 273 AA.  
 ID Q94IF6  
 AC Q94IF6  
 DT 01-DEC-2001 (TREMELrel. 19, Created)  
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)  
 DE MYB transcription factor Atmyb2.  
 GN ATMYB2.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-IN-0;  
 RA Kamiya T., Kawabe A., Miyashita N.T.;  
 RT "Nucleotide polymorphism at the Atmyb2 locus of the wild plant Arabidopsis thaliana."  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB052239; BAB62119.1; -  
 DR InterPro; IPR001395; Aldo/ket\_red.  
 DR InterPro; IPR001005; Myb\_DNA-binding.  
 DR Pfam; PF00249; myb\_DNA-binding; 2.  
 DR PROSITE; PS00063; ALDOKETO\_REDUCTASE\_3; UNKNOWN\_1.  
 DR PROSITE; PS00037; MYB\_1; UNKNOWN\_1.  
 DR PROSITE; PS00334; MYB\_2; UNKNOWN\_1.  
 DR PROSITE; PS50090; MYB\_3; 2.  
 SQ SEQUENCE 273 AA; 31433 MW; 726186432287049E CRC64;

Query Match 77.4%; Score 41; DB 10; Length 273;  
 Best Local Similarity 66.7%; Pred. No. 40;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
 :||| |||:  
 Db 22 KGPWTEED 30

## RESULT 7

Q93VZ0 PRELIMINARY; PRT; 273 AA.  
 ID Q93VZ0  
 AC Q93VZ0  
 DT 01-DEC-2001 (TREMELrel. 19, Created)  
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)  
 DE MYB transcription factor Atmyb2.  
 GN ATMYB2.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VARIOUS STRAINS;  
 RA Kamiya T., Kawabe A., Miyashita N.T.;  
 RT "Nucleotide polymorphism at the Atmyb2 locus of the wild plant Arabidopsis thaliana."  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB052250; BAB62130.1; -  
 DR EMBL; AB052232; BAB62112.1; -  
 DR EMBL; AB052233; BAB62113.1; -  
 DR EMBL; AB052236; BAB62116.1; -  
 DR EMBL; AB052237; BAB62117.1; -  
 DR EMBL; AB052240; BAB62120.1; -  
 DR EMBL; AB052242; BAB62122.1; -  
 DR EMBL; AB052243; BAB62123.1; -  
 DR EMBL; AB052245; BAB62125.1; -  
 DR EMBL; AB052246; BAB62126.1; -  
 DR EMBL; AB052247; BAB62127.1; -  
 DR InterPro; IPR001395; Aldo/ket\_red.  
 DR InterPro; IPR001005; Myb\_DNA-binding.  
 DR Pfam; PF00249; myb\_DNA-binding; 2.  
 DR PROSITE; PS00063; ALDOKETO\_REDUCTASE\_3; UNKNOWN\_1.  
 DR PROSITE; PS00037; MYB\_1; UNKNOWN\_1.  
 DR PROSITE; PS00334; MYB\_2; UNKNOWN\_1.  
 DR PROSITE; PS50090; MYB\_3; 2.  
 SQ SEQUENCE 273 AA; 31434 MW; 726F88A3C287049E CRC64;

Query Match 77.4%; Score 41; DB 10; Length 273;  
 Best Local Similarity 66.7%; Pred. No. 40;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
 :||| |||:  
 Db 22 KGPWTEED 30

## RESULT 8

Q93V46 PRELIMINARY; PRT; 273 AA.  
 ID Q93V46  
 AC Q93V46  
 DT 01-DEC-2001 (TREMELrel. 19, Created)  
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)  
 DE MYB transcription factor Atmyb2.  
 GN ATMYB2.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WS-0, KAS-1, AND OST-0;  
 RA Kamiya T., Kawabe A., Miyashita N.T.;  
 RT "Nucleotide polymorphism at the Atmyb2 locus of the wild plant Arabidopsis thaliana."  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB052249; BAB62129.1; -  
 DR EMBL; AB052241; BAB62121.1; -  
 DR EMBL; AB052244; BAB62124.1; -  
 DR InterPro; IPR001395; Aldo/ket\_red.  
 DR InterPro; IPR001005; Myb\_DNA-binding.  
 DR Pfam; PF00249; myb\_DNA-binding; 2.  
 DR PROSITE; PS00063; ALDOKETO\_REDUCTASE\_3; UNKNOWN\_1.  
 DR PROSITE; PS00037; MYB\_1; UNKNOWN\_1.  
 DR PROSITE; PS00334; MYB\_2; UNKNOWN\_1.  
 DR PROSITE; PS50090; MYB\_3; 2.  
 SQ SEQUENCE 273 AA; 31461 MW; F8472A327B97FA7E CRC64;

Query Match 77.4%; Score 41; DB 10; Length 273;  
 Best Local Similarity 66.7%; Pred. No. 40;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 EGPWLEEEE 9
DB 22 KGPWTEED 30

RESULT 9
Q39028 PRELIMINARY; PRT; 273 AA.
AC Q39028;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE ATMYB2 (MYB transcription factor ATMVB2).
GN MYB2 OR T08113.3 OR ATMVB2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=94146551; PubMed=8312738;
RA Urao T., Yamaguchi-Shinozaki K., Urao S., Shinozaki K.;
RT "An Arabidopsis myb homolog is induced by dehydration stress and its
RT gene product binds to the conserved MYB recognition sequence.";
RL Plant Cell 5:1529-1539(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA Spriggs T.A., Mason T.M., Kerlavage A.R., Adams M.D., Somerville C.R.,
RA Venter J.C.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=UK-2;
RA Kamiya T., Kawabe A., Miyashita N.T.;
RT "Nucleotide polymorphism at the Atmyb2 locus of the wild plant
RT Arabidopsis thaliana.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; D14712; BAA03534.1; -.
DR EMBL; AC002337; AAB63819.1; -.
DR EMBL; AB052248; BAB62128.1; -.
DR HSSP; P06876; IMSE.
DR TRANSFAC; T02536; -.
DR InterPro; IPR001395; Aldo/ket_red.
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF00249; myb_DNA-binding; 2.
DR SMART; SM00395; SANT; 2.
DR PROSITE; PS00063; ALDO_KETO_REDUCTASE_3; UNKNOWN_1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR PROSITE; PS00334; MYB_2; 1.
DR PROSITE; PS50090; MYB_3; 2.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 273 AA; 31448 MW; 94BAE5F38C3854DF CRC64;

Query Match 77.4%; Score 41; DB 10; Length 273;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
DB 22 KGPWTEED 30

RESULT 10
O49020 PRELIMINARY; PRT; 293 AA.
ID O49020
AC O49020

QY 1 EGPWLEEEE 9
DB 22 KGPWTEED 30

Query Match 77.4%; Score 41; DB 10; Length 273;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
DB 22 KGPWTEED 31

Query Match 77.4%; Score 41; DB 10; Length 293;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE MYB-like DNA-binding domain protein.
GN CMY-N.
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Malvales; Malvaceae; Gossypium.
OX NCBI_TaxID=3635;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ACALA SJ-2; TISSUE=OVULE;
RA Loquerio L.L., Zhang J., Wilkins T.A.;
RT "Structure and expression of six classes of myb-domain genes in
RT allotetraploid cotton (Gossypium hirsutum L.).";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 MYB-LIKE DOMAINS.
DR EMBL; AF034133; AAC04719.1; -.
DR HSSP; P06876; IMSE.
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF00249; myb_DNA-binding; 2.
DR SMART; SM00395; SANT; 2.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR PROSITE; PS00334; MYB_2; 1.
DR PROSITE; PS50090; MYB_3; 2.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 293 AA; 33746 MW; 06901890BD174772 CRC64;

Query Match 77.4%; Score 41; DB 10; Length 293;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
DB 22 KGPWTEED 30

RESULT 11
Q94JN5 PRELIMINARY; PRT; 293 AA.
ID Q94JN5
AC Q94JN5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Myb-like transcription factor Myb 5.
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Malvales; Malvaceae; Gossypium.
OX NCBI_TaxID=3635;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDRACE 'PALMERI';
RA Cronn R., Cedroni M., Haselkorn T., Osborne C., Wendel J.F.;
RT "PCR-mediated recombination in a polyploid plant.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF377316; AAK57698.1; -.
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF00249; myb_DNA-binding; 2.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR PROSITE; PS00334; MYB_2; UNKNOWN_1.
DR PROSITE; PS50090; MYB_3; 2.
SQ SEQUENCE 293 AA; 33557 MW; 4F8912A79659F53E CRC64;

Query Match 77.4%; Score 41; DB 10; Length 293;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
DB 23 KGPWTEED 31

Query Match 77.4%; Score 41; DB 10; Length 293;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```
RESULT 12
Q94JN4
ID Q94JN4 PRELIMINARY; PRT; 293 AA.
AC Q94JN4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE Myb-like transcription factor Myb 5.
DE Myb-like transcription factor Myb 5.
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Malvales; Malvaceae; Gossypium.
OX NCBI_TaxID=3635;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDRACE 'PALMERI';
RA Cronn R., Cedroni M., Haselkorn T., Osborne C., Wendel J.F.;
RT "PCR-mediated recombination in a polyploid plant.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF377317; AAK57699.1; -.
DR InterPro: IPR001005; Myb_DNA_binding.
DR Pfam: PF00249; myb_DNA-binding; 2.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR PROSITE; PS00334; MYB_2; UNKNOWN_1.
DR PROSITE; PS50090; MYB_3; 2.
DR PROSITE; PS50090; MYB_3; 2.
SQ SEQUENCE 293 AA; 33407 MW; C1FLADBA5966401A CRC64;

Query Match 77.8%; Score 41; DB 10; Length 293;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
DB 23 KGPWTEED 31

RESULT 13
P81394
ID P81394 PRELIMINARY; PRT; 268 AA.
AC P81394;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE MYB-related protein 315.
GN MYB 315
OS Antirrhinum majus (Garden snapdragon).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Veronicaceae; Antirrhinum.
OX NCBI_TaxID=4151;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JI:522; TISSUE=FLOWER BUDS;
RX MEDLINE=93005689; PubMed=1840903;
RA Jackson D., Cullane-Macia F., Prescott A.G., Roberts K., Martin C.;
RT "Expression patterns of myb genes from Antirrhinum flowers.";
RL Plant Cell 3:115-125(1991).
CC -!- FUNCTION: MAY BE A TRANSCRIPTIONAL ACTIVATOR.
CC -!- TISSUE SPECIFICITY: ROOT, STEM, LEAF, SEED POD, IMMATURE FLOWER
CC -!- SIMILARITY: BELONGS TO THE MYB FAMILY OF PROTEINS.
CC PIR; JQ061; JQ0961.
DR HSSP; O03237; 1A5J.
DR TRANSFAC; T02927; -.
DR InterPro: IPR001005; Myb_DNA_binding.
DR Pfam; PF00249; myb_DNA-binding; 2.
DR SMART; SM00395; SANT; 2.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR PROSITE; PS00334; MYB_2; 1.
DR PROSITE; PS50090; MYB_3; 2.
KW Nuclear protein; DNA-binding; Repeat; Transcription regulation.

FT DNA_BIND 9 61 MYB.
FT DNA_BIND 62 102 MYB.
SQ SEQUENCE 268 AA; 31110 MW; 886C04A4F45D26D1 CRC64;

Query Match 75.5%; Score 40; DB 10; Length 268;
Best Local Similarity 75.0%; Pred. No. 58;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GPWLEEEE 9
DB 15 GPWTEED 22

RESULT 14
Q9EJP2
ID Q9EJP2 PRELIMINARY; PRT; 310 AA.
AC Q9EJP2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Transcription factor-like protein.
DE Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98403884; PubMed=9734815;
RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
RT Sequence features of the regions of 1,367,185 bp covered by 19
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:203-216(1998).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 2 MYB-LIKE DOMAINS.
DR EMBL; AB013395; BABL1659.1; -.
DR HSSP; O03237; 1A5J.
DR InterPro: IPR001005; Myb_DNA_binding.
DR Pfam; PF00249; myb_DNA-binding; 2.
DR SMART; SM00395; SANT; 2.
DR PROSITE; PS00334; MYB_2; 1.
DR PROSITE; PS50090; MYB_3; 2.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 310 AA; 35136 MW; 9070DD86A4C2AF9D CRC64;

Query Match 75.5%; Score 40; DB 10; Length 310;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
DB 14 KGPWLPED 22

RESULT 15
Q9CQD5
ID Q9CQD5 PRELIMINARY; PRT; 311 AA.
AC Q9CQD5;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 4930413022R1K protein.
GN 4930413022R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS, AND EMBRYO;
```

```

RX MEDLINE-21085660; PubMed-11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK015133; BAB29721.1;|.
DR EMBL; AK013065; BAB28630.1; -.
DR MGD; MGI:1915236; 4930413022Rik.
SQ SEQUENCE 311 AA; 35830 MW; 9B571209DB5E91C5 CRC64;

```

```

Query Match      75.5%; Score 40; DB 11; Length 311;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 EGFWLEEE 9
   ||||| :|
Db 191 EGFWLSRQE 199

```

```

Search completed: February 5, 2003, 09:49:14
Job time : 22 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:48:26 : Search time 117 seconds

(without alignments)  
49.595 Million cell updates/sec

Title: US-09-700-329-1

Perfect score: 53

Sequence: 1 EGPWLEEE 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*

- 1: /cgn2\_6/ptodata/1/paa/PCTUS COMB.pcp.\*
- 2: /cgn2\_6/ptodata/1/paa/US06 COMB.pcp.\*
- 3: /cgn2\_6/ptodata/1/paa/US07 COMB.pcp.\*
- 4: /cgn2\_6/ptodata/1/paa/US08 COMB.pcp.\*
- 5: /cgn2\_6/ptodata/1/paa/US081 COMB.pcp.\*
- 6: /cgn2\_6/ptodata/1/paa/US082 COMB.pcp.\*
- 7: /cgn2\_6/ptodata/1/paa/US083 COMB.pcp.\*
- 8: /cgn2\_6/ptodata/1/paa/US084 COMB.pcp.\*
- 9: /cgn2\_6/ptodata/1/paa/US085 COMB.pcp.\*
- 10: /cgn2\_6/ptodata/1/paa/US086 COMB.pcp.\*
- 11: /cgn2\_6/ptodata/1/paa/US087 COMB.pcp.\*
- 12: /cgn2\_6/ptodata/1/paa/US088 COMB.pcp.\*
- 13: /cgn2\_6/ptodata/1/paa/US089 COMB.pcp.\*
- 14: /cgn2\_6/ptodata/1/paa/US090 COMB.pcp.\*
- 15: /cgn2\_6/ptodata/1/paa/US091 COMB.pcp.\*
- 16: /cgn2\_6/ptodata/1/paa/US092 COMB.pcp.\*
- 17: /cgn2\_6/ptodata/1/paa/US093 COMB.pcp.\*
- 18: /cgn2\_6/ptodata/1/paa/US094 COMB.pcp.\*
- 19: /cgn2\_6/ptodata/1/paa/US095 COMB.pcp.\*
- 20: /cgn2\_6/ptodata/1/paa/US096 COMB.pcp.\*
- 21: /cgn2\_6/ptodata/1/paa/US097 COMB.pcp.\*
- 22: /cgn2\_6/ptodata/1/paa/US098 COMB.pcp.\*
- 23: /cgn2\_6/ptodata/1/paa/US099 COMB.pcp.\*
- 24: /cgn2\_6/ptodata/1/paa/US100 COMB.pcp.\*
- 25: /cgn2\_6/ptodata/1/paa/US101 COMB.pcp.\*
- 26: /cgn2\_6/ptodata/1/paa/US102 COMB.pcp.\*
- 27: /cgn2\_6/ptodata/1/paa/US60 COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	9	1	PCT-US99-10734-7
2	53	100.0	9	1	PCT-US99-10751-1
3	53	100.0	9	11	US-08-798-423-7
4	53	100.0	9	21	US-09-700-329-1
5	53	100.0	9	21	US-09-700-378-7
6	53	100.0	9	21	US-09-700-402-1
					Sequence 7, Appli
					Sequence 1, Appli
					Sequence 7, Appli
					Sequence 1, Appli
					Sequence 7, Appli
					Sequence 1, Appli

7	53	100.0	12	11	US-08-798-423-1	Sequence 1, Appli
8	53	100.0	12	19	US-09-506-854-10	Sequence 10, Appl
9	53	100.0	12	24	US-10-016-306-10	Sequence 10, Appl
10	53	100.0	16	11	US-08-798-423-6	Sequence 6, Appli
11	53	100.0	17	1	PCT-US02-22821-60	Sequence 60, Appli
12	53	100.0	17	13	US-08-929-095-2	Sequence 2, Appli
13	53	100.0	17	20	US-09-657-276-422	Sequence 422, App
14	53	100.0	17	25	US-10-197-954-60	Sequence 60, Appl
15	50	94.3	12	1	PCT-US94-04832A-75	Sequence 75, Appl
16	50	94.3	12	8	US-08-488-320A-75	Sequence 75, Appl
17	50	94.3	17	1	PCT-US94-04832A-74	Sequence 74, Appl
18	50	94.3	17	8	US-08-488-320A-74	Sequence 74, Appl
19	50	94.3	18	10	US-08-634-546-10	Sequence 10, Appl
20	50	94.3	21	10	US-08-634-546-12	Sequence 12, Appl
21	50	94.3	24	27	US-60-162-247-2847	Sequence 2847, Ap
22	50	94.3	34	1	PCT-US94-04832A-69	Sequence 69, Appl
23	50	94.3	34	4	US-08-065-428-2	Sequence 2, Appli
24	50	94.3	34	8	US-08-488-320A-69	Sequence 69, Appl
25	50	94.3	34	10	US-08-634-546-9	Sequence 9, Appli
26	50	94.3	34	20	US-09-657-276-423	Sequence 423, App
27	50	94.3	34	25	US-10-104-607B-5	Sequence 5, Appli
28	50	94.3	34	27	US-60-389-987-196	Sequence 196, App
29	50	94.3	34	27	US-60-412-418-196	Sequence 196, App
30	50	94.3	35	10	US-08-634-546-8	Sequence 8, Appli
31	50	94.3	47	1	PCT-US94-04832A-100	Sequence 100, App
32	50	94.3	47	8	US-08-488-320A-100	Sequence 100, App
33	50	94.3	52	1	PCT-US94-04832A-99	Sequence 99, Appl
34	50	94.3	52	8	US-08-488-320A-99	Sequence 99, Appl
35	50	94.3	69	1	PCT-US94-04832A-95	Sequence 95, Appl
36	50	94.3	69	8	US-08-488-320A-95	Sequence 95, Appl
37	50	94.3	74	10	US-08-634-546-7	Sequence 7, Appli
38	50	94.3	74	27	US-60-195-053-1756	Sequence 1756, Ap
39	50	94.3	100	27	US-60-163-123-1073	Sequence 1073, Ap
40	48	90.6	16	1	PCT-US02-21768-4	Sequence 4, Appli
41	48	90.6	16	25	US-10-192-257-4	Sequence 4, Appli
42	48	90.6	17	25	US-10-104-607B-4	Sequence 4, Appli
43	46	86.8	18	26	US-10-227-012-2	Sequence 2, Appli
44	45	84.9	8	1	PCT-US94-04832A-76	Sequence 76, Appl
45	45	84.9	8	8	US-08-488-320A-76	Sequence 76, Appl

#### ALIGNMENTS

RESULT 1  
PCT-US99-10734-7  
; Sequence 7, Application PC/TUS9910734  
; GENERAL INFORMATION:  
; APPLICANT: Gevas, Philip C.  
; APPLICANT: Grimes, Stephen  
; APPLICANT: Karr, Stephen  
; APPLICANT: Michaeli, Dov  
; APPLICANT: Aphcon Corporation  
; TITLE OF INVENTION: Method for the Treatment of Gastroesophageal Reflux  
; TITLE OF INVENTION: Disease  
; FILE REFERENCE: ACG1PCT  
; CURRENT APPLICATION NUMBER: PCT/US99/10734  
; CURRENT FILING DATE: 1999-05-14  
; EARLIER APPLICATION NUMBER: 60/085,610  
; EARLIER FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 9  
; TYPE: PPT  
; ORGANISM: Unknown  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (1)  
; OTHER INFORMATION: pyroglutamic acid  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: human or  
; OTHER INFORMATION: synthetic peptide

## PCT-US99-10734-7

Query Match 100.0%; Score 53; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
| | | | | | | | |  
Db 1 EGPWLEEEE 9

RESULT 2  
PCT-US99-10751-1  
; Sequence 1, Application PC/TUS9910751  
; GENERAL INFORMATION:  
; APPLICANT: APHTON CORPORATION  
; TITLE OF INVENTION: Prevention and Treatment of Hypergastrinemia  
; FILE REFERENCE: 1102865-0035  
; CURRENT APPLICATION NUMBER: PCT/US99/10751  
; CURRENT FILING DATE: 1999-05-14  
; EARLIER APPLICATION NUMBER: US160/085,714  
; PRIOR FILING DATE: 1999-05-15  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: human or synthetic  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (1)  
; OTHER INFORMATION: pyroglutamic acid  
PCT-US99-10751-1

Query Match 100.0%; Score 53; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
| | | | | | | | |  
Db 1 EGPWLEEEE 9

## US-08-798-423-7

; Sequence 7, Application US/08798423  
; GENERAL INFORMATION:  
; APPLICANT: Gevas, Philip C.  
; APPLICANT: Karr, Stephen L.  
; APPLICANT: Grimes, Stephen  
; APPLICANT: Michaeli, Dov  
; APPLICANT: Watson, Susan A.  
; TITLE OF INVENTION: Immunological Methods for the Treatment of  
; TITLE OF INVENTION: Gastrointestinal Cancer  
; FILE REFERENCE: 1102865-0031  
; CURRENT APPLICATION NUMBER: US/08/798,423  
; CURRENT FILING DATE: 1997-02-07  
; PRIOR APPLICATION NUMBER: 60/011,411  
; PRIOR FILING DATE: 1996-02-08  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (1)  
; OTHER INFORMATION: Pyroglutamic acid residue  
US-08-798-423-7

Query Match 100.0%; Score 53; DB 11; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
| | | | | | | | |  
Db 1 EGPWLEEEE 9

RESULT 4  
US-09-700-329-1  
; Sequence 1, Application US/09700329  
; GENERAL INFORMATION:  
; APPLICANT: Gevas, Philip  
; APPLICANT: Grimes, Stephen  
; APPLICANT: Karr, Stephen  
; APPLICANT: Michaeli, Dov  
; APPLICANT: Watson, Susan  
; TITLE OF INVENTION: Prevention and Treatment of Hypergastrinemia  
; FILE REFERENCE: ACG2USA  
; CURRENT APPLICATION NUMBER: US/09/700,329  
; CURRENT FILING DATE: 2001-02-08  
; PRIOR APPLICATION NUMBER: PCT/US99/10751  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/085,714  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: human gastrin peptide  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: pyroglutamic acid  
US-09-700-329-1

Query Match 100.0%; Score 53; DB 21; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
| | | | | | | | |  
Db 1 EGPWLEEEE 9

## US-09-700-378-7

; Sequence 7, Application US/09700378  
; GENERAL INFORMATION:  
; APPLICANT: Gevas, Philip  
; APPLICANT: Stephen, Grimes  
; APPLICANT: Karr, Stephen  
; APPLICANT: Michaeli, Dov  
; TITLE OF INVENTION: Method for the Treatment of Gastroesophageal Reflux Disease  
; FILE REFERENCE: ACGUSA  
; CURRENT APPLICATION NUMBER: US/09/700,378  
; CURRENT FILING DATE: 2000-11-14  
; PRIOR APPLICATION NUMBER: PCT/US99/10734  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/085,610  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (1)..(1)  
; OTHER INFORMATION: pyroglutamic acid  
US-09-700-378-7

Query Match 100.0%; Score 53; DB 21; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9

RESULT 6  
US-09-700-402-1  
; Sequence 1, Application US/09700402  
; GENERAL INFORMATION:  
; APPLICANT: APHTON CORPORATION  
; TITLE OF INVENTION: Combination Therapy for the Treatment of Tumors  
; FILE REFERENCE: 1102865-0034  
; CURRENT APPLICATION NUMBER: US/09/700,402  
; CURRENT FILING DATE: 2001-05-04  
; PRIOR APPLICATION NUMBER: US 60/085,687  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: human or synthetic peptide  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (1)  
; OTHER INFORMATION: pyroglutamic acid  
US-09-700-402-1

Query Match 100.0%; Score 53; DB 21; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9

RESULT 7  
US-08-798-423-1  
; Sequence 1, Application US/08798423  
; GENERAL INFORMATION:  
; APPLICANT: Gevas, Philip C.  
; APPLICANT: Kart, Stephen L.  
; APPLICANT: Grimes, Stephen  
; APPLICANT: Michaeli, Dov  
; APPLICANT: Watson, Susan A.  
; TITLE OF INVENTION: Immunological Methods for the Treatment of  
; FILE REFERENCE: 1102865-0031  
; CURRENT APPLICATION NUMBER: US/08/798,423  
; CURRENT FILING DATE: 1997-02-07  
; PRIOR APPLICATION NUMBER: 60/011,411  
; PRIOR FILING DATE: 1996-02-08  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (1)  
; OTHER INFORMATION: Pyroglutamic acid residue  
US-08-798-423-1

Query Match 100.0%; Score 53; DB 11; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.37;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9

RESULT 8  
US-09-506-854-10  
; Sequence 10, Application US/09506854  
; GENERAL INFORMATION:  
; APPLICANT: Cheng, Jill  
; APPLICANT: Lasky, Laurence A.  
; TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR HOHBACH TEST ALBRITTON & HERBERT LLP  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/506,854  
; FILING DATE: 17-FEB-2000  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/991,953  
; FILING DATE: 16-DEC-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/652,971  
; FILING DATE: 24-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Vance, Dolly A.  
; REGISTRATION NUMBER: 39,054  
; REFERENCE/DOCKET NUMBER: A-63478-4/DAV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-506-854-10

Query Match 100.0%; Score 53; DB 19; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.37;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
Db 1 EGPWLEEEE 9

RESULT 9  
US-10-016-306-10  
; Sequence 10, Application US/10016306  
; GENERAL INFORMATION:  
; APPLICANT: Cheng, Jill  
; APPLICANT: Lasky, Laurence A.  
; TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
; PHOSPHATASE, PTP LAMBDA  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR HOHBACH TEST ALBRITTON & HERBERT LLP

STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppydisk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/016,306  
FILING DATE: 10-Dec-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/506,854  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 08/991,953  
FILING DATE: 16-DEC-1997  
APPLICATION NUMBER: US 08/652,971  
FILING DATE: 24-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Vance, Dolly A.  
REGISTRATION NUMBER: 39,054  
REFERENCE/DOCKET NUMBER: A-63478-4/DAV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-10-016-306-10

Query Match 100.0%; Score 53; DB 24; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.37; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0;

QY 1 EGPWLEEEE 9  
| | | | | | | | | |  
Db 1 EGPWLEEEE 9

## RESULT 10

US-08-798-423-6  
; Sequence 6, Application US/08798423  
; GENERAL INFORMATION:  
; APPLICANT: Gevas, Philip C.  
; APPLICANT: Karr, Stephen L.  
; APPLICANT: Grimes, Stephen  
; APPLICANT: Michaeli, Dov  
; APPLICANT: Watson, Susan A.  
; TITLE OF INVENTION: Immunological Methods for the Treatment of  
; TITLE OF INVENTION: Gastrointestinal Cancer  
; FILE REFERENCE: 1102865-0031  
; CURRENT APPLICATION NUMBER: US/08/798,423  
; CURRENT FILING DATE: 1997-02-07  
; PRIOR APPLICATION NUMBER: 60/011,411  
; PRIOR FILING DATE: 1996-02-08  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (1)

; OTHER INFORMATION: Pyroglutamic acid residue  
US-08-798-423-6

Query Match 100.0%; Score 53; DB 11; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.49; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0;

QY 1 EGPWLEEEE 9  
| | | | | | | | | |  
Db 1 EGPWLEEEE 9

## RESULT 11

PCT-US02-22821-60  
; Sequence 60, Application PC/TUS0222821  
; GENERAL INFORMATION:  
; APPLICANT: HK Pharmaceuticals, Inc.  
; APPLICANT: Koister, Hubert  
; APPLICANT: Siddigi, Suhaib  
; APPLICANT: Little, Daniel  
; TITLE OF INVENTION: Capture Compounds, Collections Thereof  
; TITLE OF INVENTION: And Methods For Analyzing The Proteome And Complex  
; TITLE OF INVENTION: Compositions  
; FILE REFERENCE: 24743-2305  
; CURRENT APPLICATION NUMBER: PCT/US02/22821  
; CURRENT FILING DATE: 2002-07-16  
; PRIOR APPLICATION NUMBER: 60/306,019  
; PRIOR FILING DATE: 2001-07-16  
; PRIOR APPLICATION NUMBER: 60/314,123  
; PRIOR FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: 60/363,433  
; PRIOR FILING DATE: 2002-03-11  
; NUMBER OF SEQ ID NOS: 149  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 60  
; LENGTH: 17  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
PCT-US02-22821-60

Query Match 100.0%; Score 53; DB 1; Length 17;  
Best Local Similarity 100.0%; Pred. No. 0.52; Mismatches 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0;

QY 1 EGPWLEEEE 9  
| | | | | | | | | |  
Db 1 EGPWLEEEE 9

## RESULT 12

US-08-929-095-2  
; Sequence 2, Application US/08929095  
; GENERAL INFORMATION:  
; APPLICANT: Sportsman, Richard  
; APPLICANT: Cairns, Nicholas  
; APPLICANT: Kauvar, Lawrence  
; APPLICANT: Villar, Hugo  
; APPLICANT: Lee, Sandra  
; TITLE OF INVENTION: COMPETITIVE BINDING FLUORESCENCE POLARIZATION ASSAY  
; FILE REFERENCE: 225502003900  
; CURRENT APPLICATION NUMBER: US/08/929,095  
; CURRENT FILING DATE: 1997-09-15  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 17  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Unknown  
US-08-929-095-2

Query Match 100.0%; Score 53; DB 13; Length 17;

Best Local Similarity 100.0%; Pred. No. 0.52; 0; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0;

QY 1 EGPWLEEEE 9  
| | | | | | | |  
Db 1 EGPWLEEEE 9

## RESULT 13

US-09-657-276-422  
; Sequence 422, Application US/09657276  
; GENERAL INFORMATION:  
; APPLICANT: ConjuChem, Inc.  
; APPLICANT: Bridon, Dominique  
; APPLICANT: Ezrin, Alan  
; APPLICANT: Milner, Peter  
; APPLICANT: Holmes, Darren  
; APPLICANT: Thibaudau, Karen  
; TITLE OF INVENTION: PROTECTION OF ENDOGENOUS THERAPEUTIC PEPTIDES FROM  
; TITLE OF INVENTION: PEPTIDASE ACTIVITY THROUGH CONJUGATION TO BLOOD  
; TITLE OF INVENTION: COMPONENTS  
; FILE REFERENCE: 2110  
; CURRENT APPLICATION NUMBER: US/09/657,276  
; CURRENT FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: 60/134,406  
; PRIOR FILING DATE: 1999-05-17  
; PRIOR APPLICATION NUMBER: 60/153,406  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: 60/159,783  
; PRIOR FILING DATE: 1999-10-18  
; NUMBER OF SEQ ID NOS: 1617  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 422  
; LENGTH: 17  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: Peptide  
US-09-657-276-422

Query Match 100.0%; Score 53; DB 20; Length 17;  
Best Local Similarity 100.0%; Pred. No. 0.52;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
| | | | | | | |  
Db 1 EGPWLEEEE 9

## RESULT 14

US-10-197-954-60  
; Sequence 60, Application US/10197954  
; GENERAL INFORMATION:  
; APPLICANT: K'ster, Hubert  
; APPLICANT: Siddiqui, Suhailb  
; APPLICANT: Little, Daniel  
; TITLE OF INVENTION: Capture Compounds, Collections Thereof  
; TITLE OF INVENTION: And Methods For Analyzing The Proteome And Complex  
; TITLE OF INVENTION: Compositions  
; FILE REFERENCE: 24743-2305  
; CURRENT APPLICATION NUMBER: US/10/197,954  
; CURRENT FILING DATE: 2002-07-16  
; PRIOR APPLICATION NUMBER: 60/306,019  
; PRIOR FILING DATE: 2001-07-16  
; PRIOR APPLICATION NUMBER: 60/314,123  
; PRIOR FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: 60/363,433  
; PRIOR FILING DATE: 2002-03-11  
; NUMBER OF SEQ ID NOS: 149  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 60  
; LENGTH: 17

; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-197-954-60

Query Match 100.0%; Score 53; DB 25; Length 17;  
Best Local Similarity 100.0%; Pred. No. 0.52;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
| | | | | | | |  
Db 1 EGPWLEEEE 9

## RESULT 15

PCT-US94-04832A-75  
; Sequence 75, Application PC/TUS9404832A  
; GENERAL INFORMATION:  
; APPLICANT: Ladd, Anna  
; APPLICANT: Wang, Chang Yi  
; APPLICANT: Zamb, Timothy  
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE:  
; STREET: 400 Garden City Plaza  
; CITY: Garden City  
; STATE: NY  
; COUNTRY: US  
; ZIP: 11530  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/04832A  
; FILING DATE: 13-APR-1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME:  
; REGISTRATION NUMBER:  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (516)742-4343  
; TELEFAX: (516)742-4366  
; INFORMATION FOR SEQ ID NO: 75:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
PCT-US94-04832A-75

Query Match 94.3%; Score 50; DB 1; Length 12;  
Best Local Similarity 88.9%; Pred. No. 1.1;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
| | | | | | | |  
Db 1 QGPWLEEEE 9

Search completed: February 5, 2003, 09:55:39  
Job time : 118 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:48:01 ; Search time 14.5714 Seconds  
(without alignments)  
50.530 Million cell updates/sec

Title: US-09-700-329-1  
Perfect score: 53  
Sequence: 1 EGPWLEEEE 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412725 seqs, 81809943 residues

Total number of hits satisfying chosen parameters:	412725
--	--------

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : Pending_Patents_AA_New: *
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pcp:*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pcp:*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pcp:*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pcp:*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pcp:*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pcp:*
7: /cgn2_6/ptodata/1/paa/US11_NEW_COMB.pcp:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	53	100.0	9	6	US-10-314-057-7	Sequence 7, Appli
2	45	84.9	238	6	US-10-302-267-140	Sequence 140, App
3	40	75.5	349	6	US-10-278-173-138	Sequence 138, App
4	40	75.5	349	6	US-10-278-536-114	Sequence 114, App
5	40	75.5	349	6	US-10-278-536A-114	Sequence 114, App
6	39	73.6	929	6	US-10-218-140-1980	Sequence 1980, Ap
7	39	73.6	1413	6	US-10-288-798-24	Sequence 24, Appli
8	38	71.7	12	6	US-10-314-057-1	Sequence 1, Appli
9	37	69.8	173	5	US-09-724-676-86993	Sequence 86993, A
10	37	69.8	173	5	US-09-724-676A-86993	Sequence 86993, A
11	37	69.8	367	6	US-10-278-536-206	Sequence 206, App
12	37	69.8	367	6	US-10-278-536A-206	Sequence 206, App
13	37	69.8	493	5	US-09-724-676-86996	Sequence 86996, A
14	37	69.8	493	5	US-09-724-676A-86996	Sequence 86996, A
15	37	69.8	538	5	US-09-724-676-86994	Sequence 86994, A
16	37	69.8	538	5	US-09-724-676A-86994	Sequence 86994, A
17	37	69.8	566	7	US-60-428-082-18	Sequence 18, Appli
18	37	69.8	581	5	US-09-724-676-56304	Sequence 56304, A
19	37	69.8	581	5	US-09-724-676A-56304	Sequence 56304, A
20	37	69.8	594	5	US-09-724-676-56305	Sequence 56305, A
21	37	69.8	594	5	US-09-724-676A-56305	Sequence 56305, A
22	37	69.8	707	5	US-09-724-676-86995	Sequence 86995, A
23	37	69.8	707	5	US-09-724-676A-86995	Sequence 86995, A
24	37	69.8	729	5	US-09-724-676-51589	Sequence 51589, A
25	37	69.8	729	5	US-09-724-676-51590	Sequence 51590, A
26	37	69.8	729	5	US-09-724-676-51591	Sequence 51591, A

## ALIGNMENTS

## RESULT 1

```

US-10-314-057-7
; Sequence 7, Application US/10314057
; GENERAL INFORMATION:
; APPLICANT: Gevas, Philip
; APPLICANT: Stephen, Grimes
; APPLICANT: Karr, Stephen
; APPLICANT: Michaeli, Dov
; TITLE OF INVENTION: Method for the Treatment of
; FILE REFERENCE: AC1GUSA
; CURRENT APPLICATION NUMBER: US/10/314,057
; CURRENT FILING DATE: 2002-12-06
; PRIOR APPLICATION NUMBER: US/09/700,378
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: PCT/US99/10734
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/085,610
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 9
; TYPE: PRT
; ORGANISM: human
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: pyroglutamic acid
US-10-314-057-7

```

Query Match	100.0%	Score 53;	DB 6;	Length 9;
Best Local Similarity	100.0%	Pred. No. 3.4e+05;		
Matches	9;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;

Qy 1 EGPWLEEE 9  
1 11111111  
Db 1 EGPWLEEE 9

## RESULT. 2

```

US-302-267-140
: Sequence 140, Application US/10302267
:
: GENERAL INFORMATION:
:
: APPLICANT: Keddie, James
: APPLICANT: Fromm, Michael
: APPLICANT: Heard, Jacqueline
: APPLICANT: Riechmann, Jose Luis
: APPLICANT: Adam, Luc
: APPLICANT: Broun, Pierre

```

; APPLICANT: Pineda, Omaira  
; APPLICANT: Reuber, Lynne  
; APPLICANT: Zhang, James  
; APPLICANT: Yu, Guo-Liang  
; APPLICANT: Jiang, Cai-Zhong  
; APPLICANT: Samaha, Raymond  
; APPLICANT: Pilgrim, Marsha  
; APPLICANT: Creelman, Robert  
; TITLE OF INVENTION: PLANT GENE SEQUENCES II  
; FILE REFERENCE: MBI-0007  
; CURRENT APPLICATION NUMBER: US/10/302,267  
; CURRENT FILING DATE: 2002-11-22  
; PRIOR APPLICATION NUMBER: US/09/506,720  
; PRIOR FILING DATE: 2000-02-17  
; PRIOR APPLICATION NUMBER: 60/120,880  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: 60/121,037  
; PRIOR FILING DATE: 1999-02-22  
; PRIOR APPLICATION NUMBER: 60/124,278  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: 60/129,450  
; PRIOR FILING DATE: 1999-04-15  
; PRIOR APPLICATION NUMBER: 60/135,134  
; PRIOR FILING DATE: 1999-05-20  
; PRIOR APPLICATION NUMBER: 60/144,153  
; PRIOR FILING DATE: 1999-07-15  
; PRIOR APPLICATION NUMBER: 60/161,143  
; PRIOR FILING DATE: 1999-10-22  
; PRIOR APPLICATION NUMBER: 60/162,656  
; PRIOR FILING DATE: 1999-11-01  
; NUMBER OF SEQ ID NOS: 218  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 140  
; LENGTH: 238  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: G1311  
US-10-302-267-140

Query Match 84.9%; Score 45; DB 6; Length 238;  
Best Local Similarity 87.5%; Pred. No. 4;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPWLEEEE 9  
|||||  
Db 12 GPWLEED 19

RESULT 3  
US-10-278-173-138  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Cai-Zhong  
; APPLICANT: Broun, Pierre  
; APPLICANT: Riechmann, Jose-Luis  
; APPLICANT: Pineda, Omaira  
; APPLICANT: Zhang, James  
; APPLICANT: Yu, Guo-Liang  
; APPLICANT: Pilgrim, Marsha  
; APPLICANT: Keddle, James  
; APPLICANT: Heard, Jacqueline  
; APPLICANT: Reuber, Lynne  
; APPLICANT: Ratcliffe, Oliver  
; APPLICANT: Adam, Luc  
; APPLICANT: Samaha, Raymond  
; TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION  
; FILE REFERENCE: MBI-009  
; CURRENT APPLICATION NUMBER: US/10/278,173  
; CURRENT FILING DATE: 2002-10-21  
; PRIOR APPLICATION NUMBER: US/09/533,392  
; PRIOR FILING DATE: 2000-03-22  
; PRIOR APPLICATION NUMBER: 60/125,814

; PRIOR FILING DATE: 1999-03-23  
; NUMBER OF SEQ ID NOS: 177  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 138  
; LENGTH: 349  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: G220  
US-10-278-173-138

Query Match 75.5%; Score 40; DB 6; Length 349;  
Best Local Similarity 66.7%; Pred. No. 38;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
:|||||  
Db 15 KGPWLPEED 23

RESULT 4  
US-10-278-536-114  
; Sequence 114, Application US/10278536  
; GENERAL INFORMATION:  
; APPLICANT: Samaha, Raymond  
; APPLICANT: Heard, Jacqueline  
; APPLICANT: Jiang, Cai-Zhong  
; APPLICANT: Pineda, Omaira  
; APPLICANT: Reuber, Lynne  
; APPLICANT: Riechmann, Jose-Luis  
; APPLICANT: Yu, Guo-Liang  
; APPLICANT: Keddle, James  
; APPLICANT: Ratcliffe, Oliver  
; APPLICANT: Pilgrim, Marsha  
; APPLICANT: Adam, Luc  
; APPLICANT: Broun, Pierre  
; TITLE OF INVENTION: STRESS-INDUCED POLYNUCLEOTIDES  
; FILE REFERENCE: MBI-011  
; CURRENT APPLICATION NUMBER: US/10/278,536  
; CURRENT FILING DATE: 2002-10-22  
; PRIOR APPLICATION NUMBER: 60/125,814  
; PRIOR FILING DATE: 1999-03-23  
; NUMBER OF SEQ ID NOS: 238  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 114  
; LENGTH: 349  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: G220  
US-10-278-536-114

Query Match 75.5%; Score 40; DB 6; Length 349;  
Best Local Similarity 66.7%; Pred. No. 38;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9  
:|||||  
Db 15 KGPWLPEED 23

RESULT 5  
US-10-278-536A-114  
; Sequence 114, Application US/10278536A  
; GENERAL INFORMATION:  
; APPLICANT: Samaha, Raymond  
; APPLICANT: Heard, Jacqueline  
; APPLICANT: Jiang, Cai-Zhong  
; APPLICANT: Pineda, Omaira  
; APPLICANT: Reuber, Lynne  
; APPLICANT: Riechmann, Jose-Luis  
; APPLICANT: Yu, Guo-Liang  
; APPLICANT: Keddle, James

```

; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; FILE REFERENCE: MBI-011
; TITLE OF INVENTION: STRESS-INDUCED POLYNUCLEOTIDES
; CURRENT APPLICATION NUMBER: US/10/278,536A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 114
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G220
US-10-278-536A-114

Query Match      75.5%; Score 40; DB 6; Length 349;
Best Local Similarity 66.7%; Pred. No. 38;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
Db 15 KGPWLPEED 23

RESULT 6
US-10-218-140-1980
; Sequence 1980, Application US/10218140
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 15966-543 CON
; CURRENT APPLICATION NUMBER: US/10/218,140
; CURRENT FILING DATE: 2002-08-12
; PRIOR APPLICATION NUMBER: 09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/127,728
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/127,636
; PRIOR FILING DATE: 1999-04-02
; PRIOR APPLICATION NUMBER: 60/127,607
; PRIOR FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 6322
; SOFTWARE: Curator Version 1.0
; SEQ ID NO 1980
; LENGTH: 929
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-218-140-1980

Query Match      73.6%; Score 39; DB 6; Length 929;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GPWLEEEE 9
Db 171 GPWAEEEK 178

RESULT 7
US-10-288-798-24
; Sequence 24, Application US/10288798
; GENERAL INFORMATION:
; APPLICANT: BANDMAN, Olga; NGUYEN, Dannie B.;
; APPLICANT: WALIA, Narinder K.; HAFALIA, April J.A.;
; APPLICANT: YAO, Monique G.; GANDHI, Ameena R.;
; APPLICANT: GURURAJAN, Rajagopa; DING, Li;

```

```

; APPLICANT: PATTERSON, Chandra; YUE, Henry;
; APPLICANT: BAUGHN, Mariah R.; TRIBOULEY, Catherine M.;
; APPLICANT: THORNTON, Michael; ELLIOTT, Vicki S.;
; APPLICANT: LU, Yan; ISON, Craig H.; Y. Tom;
; APPLICANT: AU-YOUNG, Janice; TANG, Y. Tom;
; APPLICANT: AZIMZAI, Yalda; BURRILL, John D.;
; APPLICANT: MARCUS, Gregory A.; ZINGLER, Kurt A.;
; APPLICANT: LU, Dyung Aina M.; LAL, Preeti G.;
; APPLICANT: RAMKUMAR, Jayalaxmi; WARREN, Bridget A.;
; APPLICANT: KEARNEY, Liam; POLICKY, Jennifer L.;
; APPLICANT: THANGAVELU, Kavitha; BURFORD, Neil
; TITLE OF INVENTION: HUMAN KINASES
; FILE REFERENCE: PI-0209 USA
; CURRENT APPLICATION NUMBER: US/10/288,798
; CURRENT FILING DATE: 2002-11-01
; PRIOR APPLICATION NUMBER: PCT/US01/27219
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/240,542
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/238,389
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 60/236,499
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: US 60/234,902
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US 60/232,654
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: US 60/231,357
; PRIOR APPLICATION NUMBER: US 60/229,873
; PRIOR FILING DATE: 2000-08-31
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PERL Program
; SEQ ID NO 24
; LENGTH: 1413
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc:feature
; OTHER INFORMATION: Incyte ID No: 7477484CD1
US-10-288-798-24

Query Match      73.6%; Score 39; DB 6; Length 1413;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 PWLEEEE 9
Db 636 PWVEEEE 642

RESULT 8
US-10-314-057-1
; Sequence 1, Application US/10314057
; GENERAL INFORMATION:
; APPLICANT: Gevas, Philip
; APPLICANT: Stephen, Grimes
; APPLICANT: Karr, Stephen
; APPLICANT: Michael, Dov
; TITLE OF INVENTION: Method for the Treatment of Gastroesophageal Reflux Disease
; FILE REFERENCE: ACGIUSA
; CURRENT APPLICATION NUMBER: US/10/314,057
; CURRENT FILING DATE: 2002-12-06
; PRIOR APPLICATION NUMBER: US/09/700,378
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: PCT/US99/10734
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/085,610
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 12

```

```
; TYPE: PRT
; ORGANISM: human
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)..(1)
; OTHER INFORMATION: pyroglutamic acid
US-10-314-057-1

Query Match          71.7%; Score 38; DB 6; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EGPWLE 6
Db 1 EGPWLE 6

RESULT 9
US-09-724-676-86993
; Sequence 86993, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 86993
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-86993

Query Match          69.8%; Score 37; DB 5; Length 173;
Best Local Similarity 55.6%; Pred. No. 62;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
Db 87 KGPWTKEED 95

RESULT 10
US-09-724-676A-86993
; Sequence 86993, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 86993
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-86993

Query Match          69.8%; Score 37; DB 5; Length 173;
Best Local Similarity 55.6%; Pred. No. 62;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
Db 87 KGPWTKEED 95

RESULT 11
US-10-278-536-206
; Sequence 206, Application US/10278536
; GENERAL INFORMATION:
```

```
; APPLICANT: Samaha, Raymond
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Keddle, James
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; TITLE OF INVENTION: STRESS-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-011
; CURRENT APPLICATION NUMBER: US/10/278,536
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 206
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G216
US-10-278-536-206

Query Match          69.8%; Score 37; DB 6; Length 367;
Best Local Similarity 55.6%; Pred. No. 1.2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EGPWLEEEE 9
Db 50 KGPWLPEQD 58

RESULT 12
US-10-278-536A-206
; Sequence 206, Application US/10278536A
; GENERAL INFORMATION:
; APPLICANT: Samaha, Raymond
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Keddle, James
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; TITLE OF INVENTION: STRESS-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-011
; CURRENT APPLICATION NUMBER: US/10/278,536A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 206
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G216
US-10-278-536A-206

Query Match          69.8%; Score 37; DB 6; Length 367;
Best Local Similarity 55.6%; Pred. No. 1.2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

Qy 1 EGPWLEEEE 9  
:|||||:|:  
Db 50 KGPWLPEQD 58

## RESULT 13

US-09-724-676-86996  
; Sequence 86996, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 86996  
; LENGTH: 493  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-86996

Query Match 69.8%; Score 37; DB 5; Length 493;  
Best Local Similarity 55.6%; Pred. No. 1.6e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
:|||||:|:  
Db 87 KGPWTKEED 95

## RESULT 14

US-09-724-676A-86996  
; Sequence 86996, Application US/09724676A  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676A  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 86996  
; LENGTH: 493  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676A-86996

Query Match 69.8%; Score 37; DB 5; Length 493;  
Best Local Similarity 55.6%; Pred. No. 1.6e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
:|||||:|:  
Db 87 KGPWTKEED 95

## RESULT 15

US-09-724-676-86994  
; Sequence 86994, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 86994  
; LENGTH: 538  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-86994

Query Match 69.8%; Score 37; DB 5; Length 538;  
Best Local Similarity 55.6%; Pred. No. 1.8e+02;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EGPWLEEEE 9  
:|||||:|:  
Db 87 KGPWTKEED 95

Search completed: February 5, 2003, 09:50:58  
Job time : 14.5714 secs



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:40:45 ; Search time 35.4286 Seconds  
(without alignments)  
45.133 Million cell updates/sec

Title: US-09-700-329-2

Perfect score: 75

Sequence: 1 ELGPQGRPPPC 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	100.0	12	11 AAR06335	Antigenic peptide
2	75	100.0	12	21 AAY51306	Human gastrin G34
3	75	100.0	12	21 AAY51313	Human gastrin G34
4	75	100.0	12	21 AAY49307	Human gastrin G34
5	56	74.7	11	15 AAY62738	Human gastrin G34
6	55	73.3	139	22 AAU30337	Novel human secret
7	55	73.3	169	22 AAU30339	Novel human secret
8	50	66.7	112	22 AAO07572	Human polypeptide
9	50	66.7	129	22 AAO07572	Human polypeptide
10	49	65.3	11	11 AAR06247	Antigenic peptide

11	49	65.3	11	21	AAY51310	Human gastrin G17
12	49	65.3	11	21	AAY49304	Human gastrin G17
13	49	65.3	146	22	AAO0530	Human polypeptide
14	49	65.3	383	21	AAB38442	Fragment of human
15	49	65.3	383	22	AAW25613	Human protein sequ
16	48	64.0	16	16	AAW78285	GnRH immunomimic a
17	48	64.0	16	21	AAY58141	Conadotropin relea
18	48	64.0	16	23	AAW66081	Conadotropin relea
19	48	64.0	102	21	AAW42786	Human ORFX ORF2550
20	47	62.7	118	22	AAW48788	Propionibacterium
21	47	62.7	362	19	AAW69717	Herpesviral VP22 p
22	47	62.7	415	13	AAW28638	UL26 protease dele
23	47	62.7	418	18	AAW26545	Mouse Huntington's
24	47	62.7	463	22	ABW59725	Drosophila melanog
25	47	62.7	603	13	AAW28641	UL26 protease dele
26	47	62.7	615	13	AAW28642	UL26 protease dele
27	47	62.7	626	13	AAW28640	UL26 protease dele
28	47	62.7	635	13	AAW28634	UL26 protease dele
29	47	62.7	635	13	AAW28650	UL26 protease dele
30	47	62.7	635	13	AAW28649	UL26 protease dele
31	47	62.7	635	13	AAW28651	UL26 protease dele
32	47	62.7	635	13	AAW28652	UL26 protease dele
33	47	62.7	635	13	AAW28647	UL26 protease dele
34	47	62.7	635	13	AAW28648	UL26 protease dele
35	47	62.7	635	16	AAW71031	HSV-1 UL26 gene pr
36	47	62.7	636	13	AAW28635	UL26 protease inse
37	47	62.7	636	13	AAW28636	UL26 protease inse
38	47	62.7	636	13	AAW28637	UL26 protease inse
39	47	62.7	3119	18	AAW36888	Mouse Huntington's
40	47	62.7	3119	19	AAW44743	Mouse Huntington's
41	46.5	62.0	674	22	ABG09037	Novel human diagno
42	46	61.3	12	21	AAY51312	Rat gastrin G17 de
43	46	61.3	12	21	AAY49306	Human gastrin G17 de
44	46	61.3	119	22	ABG13925	Human gastrin G17 de
45	46	61.3	134	22	AAO05578	Human polypeptide

#### ALIGNMENTS

RESULT 1  
AAR06335  
ID AAR06335 standard; protein; 12 AA.  
AC AAR06335;  
XX  
XX  
DT 07-DEC-1990 (first entry)  
XX  
XX Antigenic peptide fragment selected from the 22 N-terminal AAs of tetrateseracontagastrin (G34).  
DE  
DE  
XX Gastrin; tumours; peptic ulcers; diptheria toxoid; tetanus toxin;  
XX  
XX EP380230-A.  
XX  
XX 01-AUG-1990.  
XX  
XX 17-JAN-1990; 90EP-0300456.  
XX  
XX 12-MAY-1989; 89US-0351193.  
PR  
PR 24-JAN-1989; 89US-0301353.  
XX  
XX (APHT-) APHTON CORP.  
PA  
XX  
XX Gevas PC, Grimes S, Karr SL, Littenberg RL;  
XX  
XX WPI; 1990-233029/31.  
XX  
XX Immunogens against gastrin peptide(s) - used to induce antibodies that specifically neutralise single form of gastrin, G17 or G34  
XX  
XX Claim 12; Page 19; 32pp; English.  
XX

CC Antigenic fragments may be attached to an immunogenic carrier  
 CC and used to raise Abs to a specific single form of Gastrin  
 CC ie. G17 or G34. Peptide fragments capable of binding to these  
 CC Abs are useful in neutralising anti-gastrin Abs in vivo.

XX Sequence 12 AA;

Query Match 100.0%; Score 75; DB 11; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 0.00093;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ELGPGGRPPPC 12  
 |||||  
 Db 1 ELGPGGRPPPC 12

RESULT 2

AA51306  
 ID AAY51306 standard; peptide; 12 AA.

AC AAY51306;

DT 14-APR-2000 (first entry)

DE Human gastrin G34 peptide.

XX Immunogenic; treatment; hypergastrinemia; gastrin G17; gastrin G34;  
 KW pernicious anemia; anti-ulcer; proton pump inhibitor; colorectal disease.

OS Homo sapiens.

PN WO9959631-A1.

PD 25-NOV-1999.

PF 14-MAY-1999; 99WO-US10751.

PR 15-MAY-1998; 98US-0085714.

PA (APHT-) APHTON CORP.

PI Gevas PC, Grimes S, Karr S, Michaeli D, Watson S;

XX WPI; 2000-116301/10.

XX Treating or preventing hypergastrinemia comprising administration of,  
 PT e.g. anti-gastrin antibodies -

PS Disclosure; Page 11; 44pp; English.

XX This invention describes a novel method for the treatment or preventing  
 CC hypergastrinemia by administering to a patient a gastrin G17 and/or G34  
 CC peptide fragment linked by an amino acid spacer to an immunogenic  
 CC carrier. The methods are used to treat hypergastrinemic patients,  
 CC particularly those with pernicious anemia, those receiving treatment  
 CC with anti-ulcer agents such as proton pump inhibitors (particularly  
 CC omeprazole or lansoprazole) or H<sub>2</sub> receptor blocking agents or  
 CC antagonists, or those having colorectal disorders or diseases. This  
 CC sequence represents the human gastrin G34 peptide which is used to  
 CC illustrate the method of the invention.

XX Sequence 12 AA;

Query Match 100.0%; Score 75; DB 21; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 0.00093;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ELGPGGRPPPC 12  
 |||||  
 Db 1 ELGPGGRPPPC 12

RESULT 3

AA51313  
 ID AAY51313 standard; peptide; 12 AA.  
 XX AAY51313;  
 AC AAY51313;  
 DT 14-APR-2000 (first entry)  
 DE Human gastrin G34 derived immunogen hG34(6).

XX Immunogenic; treatment; hypergastrinemia; gastrin G17; gastrin G34;  
 KW pernicious anemia; anti-ulcer; proton pump inhibitor; colorectal disease.

OS Homo sapiens.

PN WO9959631-A1.

PD 25-NOV-1999.

PF 14-MAY-1999; 99WO-US10751.

PR 15-MAY-1998; 98US-0085714.

PA (APHT-) APHTON CORP.

PI Gevas PC, Grimes S, Karr S, Michaeli D, Watson S;

XX WPI; 2000-116301/10.

XX Treating or preventing hypergastrinemia comprising administration of,  
 PT e.g. anti-gastrin antibodies -

PS Example 3; Page 18; 44pp; English.

XX This invention describes a novel method for the treatment or preventing  
 CC hypergastrinemia by administering to a patient a gastrin G17 and/or G34  
 CC peptide fragment linked by an amino acid spacer to an immunogenic  
 CC carrier. The methods are used to treat hypergastrinemic patients,  
 CC particularly those with pernicious anemia, those receiving treatment  
 CC with anti-ulcer agents such as proton pump inhibitors (particularly  
 CC omeprazole or lansoprazole) or H<sub>2</sub> receptor blocking agents or  
 CC antagonists, or those having colorectal disorders or diseases. This  
 CC sequence represents a human gastrin G34 derived immunogen which is used  
 CC to illustrate the method of the invention.

XX Sequence 12 AA;

Query Match 100.0%; Score 75; DB 21; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 0.00093;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ELGPGGRPPPC 12  
 |||||  
 Db 1 ELGPGGRPPPC 12

RESULT 4

AA51307

ID AAY49307 standard; peptide; 12 AA.

XX AAY49307;

DT 06-MAR-2000 (first entry)

XX Human tetraatriacontagastrin (G34) immunogenic peptide 5.

XX Gastroesophageal reflux disease; GERD; gastrin; anti-gastrin antibody;  
 KW histamine H<sub>2</sub>; proton pump inhibitor; acid output; stomach; therapy;  
 KW esophagitis; immunogen; human; tetraatriacontagastrin; G34.

XX Synthetic.

OS Homo sapiens.

XX Key Location/Qualifiers

```

FT Modified-site 1
FT /label= pGlu
FT /note= "pyroglutamate"
XX
PN W09959612-A1.
XX
XX 25-NOV-1999.
XX
PF 14-MAY-1999; 99WO-US10734.
XX
PR 15-MAY-1998; 98US-0085610.
XX
XX (APHT-) APHTON CORP.
XX
XX Gevas PC, Grimes S, Karr S, Michaeli D;
XX WPI; 2000-062378/05.
XX
XX Method for treatment of gastroesophageal reflux disease (GERD) -
XX
XX Example 4; Page 11; 24pp; English.
XX
XX The invention relates to the treatment of gastroesophageal reflux
XX disease (GERD) that comprises administering to a patient an immunogenic
XX composition which generates anti-gastrin antibodies, which bind to
XX gastrin, in a patient; and administering histamine H2 antagonist or a
XX proton pump inhibitor. The method provides a more effective method for
XX controlling acid output by the stomach. The therapy is less costly. High
XX gastrin levels associated with standard therapies are neutralized and
XX undesirable side effects are reduced. The method permits a reduced
XX dosage of acid reducing agent both at the acid producing level as well as
XX the acid production stimulating level (gastrin). Reduction of dosages is
XX desirable for prolonged treatment of GERD. In a combination therapy with
XX H2 agonists or proton pump inhibitors, anti-gastrin 17 antibody titers
XX can be maintained by occasional booster shots while gastric acid
XX inhibitor dosing is reduced. Immunization allows a sufficient time for
XX the esophagitis to completely heal and no surgery is required. The
XX present sequence represents a human tetratriacontagastrin (G34)
XX immunogenic peptide.
XX CC AAY49307-308 represent human tetratriacontagastrin (G34) immunogenic
XX peptides.
XX
XX Sequence 12 AA;
XX
Query Match 100.0%; Score 75; DB 21; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00093;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ELGPGGRPPPPC 12
Db :|||||
1 ELGPGGRPPPPC 12

RESULT 5
AAR62738
ID AAR62738 standard; peptide; 11 AA.
XX
XX AAR62738;
XX
XX 21-SEP-1995 (first entry)
XX
XX Gastrin haptens.
XX
XX Helper T cell epitope; universal immune stimulator; invasive; haptens;
XX gastrin; peptic ulcers; gastrin-stimulated tumours.
XX
XX Homo sapiens.
XX
XX W09425060-A.
XX
XX 10-NOV-1994.
XX
XX 28-APR-1994; 94WO-US04832.
XX

```

```

XX 27-APR-1993; 93US-0057166.
PR 14-APR-1994; 94US-0229275.
XX
XX (LADD/) LADD A E.
PA (WANG/) WANG C Y.
PA (ZAMB/) ZAMB T.
XX
PI Ladd AE, Wang CY, Zamb T;
XX
XX WPI; 1994-357910/44.
XX
XX Immunogenic luteinising hormone releasing hormone peptide(s) -
XX that suppress LHRH activity in males and females
XX
XX Claims 27, 37; Page 96; 213pp; English.
XX
XX Synthetic immunogenic peptides are provided in which a universal immune
XX stimulator is linked to a peptide or protein haptens containing B cell
XX and/or cytotoxic T lymphocyte epitopes, giving a product which causes
XX potent immune responses to the coupled peptide or protein. The
XX stimulator consists of (A) a promiscuous helper T cell epitope (Th)
XX which elicits an immune response to the coupled peptide in members of
XX a heterogeneous population expressing diverse HLA phenotypes, and (B)
XX an adjuvant peptide sequence from the invasive protein of Yersinia.
XX Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
XX invasive and Th domains and between the immune stimulator and haptens
XX components. When the haptens is LHRH, then optionally the invasive domain
XX can be omitted from the immune stimulator component.
XX The present sequence is an example of a gastrin haptens which can be
XX bound to the immune stimulator to form a vaccine for treating
XX peptic ulcer disease or gastrin-stimulated tumours.
XX
XX Sequence 11 AA;
XX
Query Match 74.7%; Score 56; DB 15; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.029;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ELGPGGRPPPP 11
Db :|||||
1 QLGPGRPPPP 11

RESULT 6
AAU30337
ID AAU30337 standard; Protein; 139 AA.
XX
XX AAU30337;
XX
XX 18-DEC-2001 (first entry)
XX
XX Novel human secreted protein #828.
XX
XX Human; vaccination; gene therapy; nutritional supplement;
XX stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
XX Homo sapiens.
XX
XX W0200179449-A2.
XX
XX 25-OCT-2001.
XX
XX 16-APR-2001; 2001WO-US08656.
XX
XX 18-APR-2000; 2000US-0552929.
XX
XX 26-JAN-2001; 2001US-0770160.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX

```



CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 112 AA;

Query Match 66.7%; Score 50; DB 22; Length 112;

Best Local Similarity 88.9%; Pred. No. 15;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GPQGRPPPP 11

DB 49 GPQGGPPPP 57

#### RESULT 9

AA000748

ID AAO00748 standard; Protein; 129 AA.

XX AC

XX AAO00748;

XX 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 14640.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

XX 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AAI80679.

XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -

PS Claim 20; SEQ ID NO 14640; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 129 AA;

Query Match 66.7%; Score 50; DB 22; Length 129;

Best Local Similarity 77.8%; Pred. No. 17;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 GPQGRPPPP 11

DB 20 GPKGKPPPP 28

#### RESULT 10

AA06247

ID AAR06247 standard; protein; 11 AA.

XX AC

XX AAR06247;

XX 07-DEC-1990 (first entry)

XX Antigenic peptide fragment selected from the 12 N-terminal AAs of  
 DE heptadecagastrin (G17).

XX Gastrin; tumours; peptic ulcers; diphtheria toxoid; tetanus toxin;

PN EP380230-A.

XX 01-AUG-1990.

XX 17-JAN-1990; 90EP-0300456.

XX 12-MAY-1989; 89US-0351193.

PR 24-JAN-1989; 89US-0301353.

XX (APHT-) APHTON CORP.

XX Gevas PC, Grimes S, Karr SL, Littenberg RL;

XX WPI; 1990-233029/31.

XX Immunogens against gastrin peptide(s) - used to induce antibodies  
 PT that specifically neutralise single form of gastrin, G17 or G34  
 PT Claim 8; Page 19; 32pp; English.

XX Antigenic fragments may be attached to an immunogenic carrier  
 CC and used to raise Abs to a specific single form of Gastrin  
 CC ie. G17 or G34. Peptide fragments capable of binding to these  
 CC Abs are useful in neutralising anti-gastrin Abs in vivo.

XX SQ Sequence 11 AA;

Query Match 65.3%; Score 49; DB 11; Length 11;

Best Local Similarity 80.0%; Pred. No. 2.5;

Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 GPQGRPPPP 12

DB 2 GPWLRRPPPP 11

#### RESULT 11

AA51310

ID AAY51310 standard; peptide; 11 AA.

XX AC

XX AAY51310;

XX 14-APR-2000 (first entry)

XX Human gastrin G17 derived immunogen hg17(5).

DE Immunogenic; treatment; hypergastrinemia; gastrin G17; gastrin G34;  
 KW pernicious anemia; anti-ulcer; proton pump inhibitor; colorectal disease.

```

XX OS Homo sapiens.
XX PN WO9959631-A1.
XX PD 25-NOV-1999.
XX PF 14-MAY-1999; 99WO-US10751.
XX PR 15-MAY-1998; 98US-0085714.
XX PA (APHT-) APHTON CORP.
XX PI Gevas PC, Grimes S, Karr S, Michaeli D, Watson S;
XX DR WPI; 2000-116301/10.
XX PT Treating or preventing hypergastrinemia comprising administration of,
XX PT e.g. anti-gastrin antibodies -
XX PS Example 3; Page 18; 44pp; English.
XX CC This invention describes a novel method for the treatment or preventing
XX CC hypergastrinemia by administering to a patient a gastrin G17 and/or G34
XX CC peptide fragment linked by an amino acid spacer to an immunogenic
XX CC carrier. The methods are used to treat hypergastrinemic patients,
XX CC particularly those with pernicious anemia, those receiving treatment
XX CC with anti-ulcer agents such as proton pump inhibitors (particularly
XX CC omeprazole or lansoprazole) or H2 receptor blocking agents or
XX CC antagonists, or those having colorectal disorders or diseases. This
XX CC sequence represents a human gastrin G17 derived immunogen which is used
XX CC to illustrate the method of the invention.
XX SQ Sequence 11 AA;
XX
Query Match 65.3%; Score 49; DB 21; Length 11;
Best Local Similarity 80.0%; Pred. No. 2.5;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GPOGRPPPPC 12
DB 11 |||||||
2 GPWLRRPPPC 11

RESULT 12
AAV49304
ID AAY49304 standard; peptide: 11 AA.
XX AC AAY49304;
XX DT 06-MAR-2000 (first entry)
XX DE Human heptadecagastrin (G17) immunogenic peptide 2.
XX KW Gastroesophageal reflux disease; GERD; gastrin; anti-gastrin antibody;
XX KW histamine H2; proton pump inhibitor; acid output; stomach; therapy;
XX KW esophagitis; immunogen; human; heptadecagastrin; G17.
XX OS Synthetic.
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT Modified-site 1
XX FT /label= pglu
XX FT /note= "pyroglutamate"
XX
XX PN WO9959612-A1.
XX PD 25-NOV-1999.
XX PF 14-MAY-1999; 99WO-US10734.
XX PR 15-MAY-1998; 98US-0085610.

```

```

XX PA (APHT-) APHTON CORP.
XX PI Gevas PC, Grimes S, Karr S, Michaeli D;
XX DR WPI; 2000-062378/05.
XX PT Method for treatment of gastroesophageal reflux disease (GERD) -
XX PS Example 4; Page 11; 24pp; English.
XX CC The invention relates to the treatment of gastroesophageal reflux
XX CC disease (GERD) that comprises administering to a patient an immunogenic
XX CC composition which generates anti-gastrin antibodies, which bind to
XX CC gastrin, in a patient; and administering histamine H2 antagonist or a
XX CC proton pump inhibitor. The method provides a more effective method for
XX CC controlling acid output by the stomach. The therapy is less costly. High
XX CC gastrin levels associated with standard therapies are neutralized and
XX CC undesirable side effects are reduced. The method permits a reduced
XX CC dosage of acid reducing agent both at the acid producing level as well as
XX CC the acid production stimulating level (gastrin). Reduction of dosages is
XX CC desirable for prolonged treatment of GERD. In a combination therapy with
XX CC H2 agonists or proton pump inhibitors, anti-gastrin 17 antibody titers
XX CC can be maintained by occasional booster shots while gastric acid
XX CC inhibitor dosing is reduced. Immunization allows a sufficient time for
XX CC the esophagitis to completely heal and no surgery is required. Sequences
XX CC AAY49303-306 represent human heptadecagastrin (G17) immunogenic
XX CC peptides.
XX SQ Sequence 11 AA;
XX
Query Match 65.3%; Score 49; DB 21; Length 11;
Best Local Similarity 80.0%; Pred. No. 2.5;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GPOGRPPPPC 12
DB 11 |||||||
2 GPWLRRPPPC 11

RESULT 13
AAO00530
ID AAO00530 standard; Protein; 146 AA.
XX AC AAO00530;
XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 14422.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200164835-A2.
XX PD 07-SEP-2001.
XX XX 26-FEB-2001; 2001WO-US04927.
XX PR 28-FEB-2000; 2000US-0515126.
XX PR 18-MAY-2000; 2000US-0577409.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-514838/56.
XX DR N-PSDB; AAI80461.
XX

```

PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PS disorders -

XX Claim 20; SEQ ID NO 14422; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 146 AA;

Query Match 65.3%; Score 49; DB 22; Length 146;  
Best Local Similarity 80.0%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 LGPQGRPPPP 11  
||| | ||||  
Db 115 LGPPGAPPPP 124

RESULT 14

AAB38442  
ID AAB38442 standard; peptide; 383 AA.

AC AAB38442;

DT 31-JAN-2001 (first entry)

DE Fragment of human secreted protein encoded by gene 19 clone HTADW91.

KW Immunosuppressive; antiarthritic; antirheumatic; antiproliferative;  
KW cytostatic; cardiant; vasotropic; cerebroprotective; neuroprotective;  
KW neutropic; antibacterial; virucide; fungicide; ophthalmological; human;  
KW vulnary; gene therapy; infection; secreted protein.

XX Homo sapiens.

XX WO200061623-A1.

PD 19-OCT-2000.

PF 06-APR-2000; 2000WO-US08979.

PR 09-APR-1999; 99US-0128693.

XX 26-APR-1999; 99US-0130991.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ni J, Komatsoulis GA, Rosen CA, Soppet DR, Shi Y;  
PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Moore PA, Birse CE;  
PI Young PE;

XX WPI; 2000-647418/62.

XX New nucleic acid molecules encoding 62 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -

PS Disclosure; Page 55; 716pp; English.

XX Sequences AAB38321-B38396 represent the amino acid sequences of 62

CC human secreted proteins encoded by the genes AAC69512-C69587. The genes  
CC and proteins are useful for preventing, ameliorating or treating medical  
CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
CC a range of human tissues disclosed in the specification. The nucleic  
CC acids, proteins, antibodies and (ant)agonists are useful in the  
CC diagnosis, treatment and prevention of: (a) autoimmune diseases e.g.  
CC rheumatoid arthritis; (b) hyperproliferative disorders e.g. neoplasms  
CC of the breast or liver; (c) cardiovascular disorders e.g. cardiac  
CC arrest; (d) cerebrovascular disorders e.g. cerebral ischemia; (e)  
CC angio genesis; (f) nervous system disorders e.g. Alzheimer's disease; (g)  
CC infections caused by bacteria, viruses and fungi; and (h) ocular  
CC disorders e.g. corneal infection. The polypeptides can also be used to  
CC aid wound healing and epithelial cell proliferation, to prevent skin  
CC aging due to sunburn, to maintain organs before transplantation, for  
CC supporting cell culture of primary tissues, to regenerate tissues and in  
CC chemotaxis.

XX Sequence 383 AA;

Query Match 65.3%; Score 49; DB 21; Length 383;  
Best Local Similarity 80.0%; Pred. No. 62;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 LGPQGRPPPP 11  
||| | ||||  
Db 5 LGPPGSPPPP 14

RESULT 15

AAM25613  
ID AAM25613 standard; Protein; 383 AA.

AC AAM25613;

DT 16-OCT-2001 (first entry)

DE Human protein sequence SEQ ID NO:1128.

KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiaggregant; haemostatic; vulnary; antiulcer; osteopathic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; pancreatitis;  
KW antianaphylactic; rheumatoid arthritis; septic shock; inflammation;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder.

XX Homo sapiens.

XX WO200153455-A2.

XX 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457603/49.

XX N-PSDB; AAH99554.

PT Isolated human polynucleotides encoding polypeptides, useful for the  
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
XX  
XX  
PS Claim 20; Page 234; 1217pp; English.  
XX  
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
CC AAM25963. The proteins can have activities based on the tissues and  
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;  
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;  
CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;  
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
CC encoding them can be used in gene therapy, antisense therapy and vaccine  
CC production. The proteins and polynucleotides are useful for screening for  
CC agonists or antagonists of a protein and for the treatment and diagnosis  
CC of disorders associated with the activity of a protein e.g. inflammation,  
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
CC neurological disorders.

XX SQ Sequence 383 AA;

Query Match 65.3%; Score 49; DB 22; Length 383;  
Best Local Similarity 80.0%; Pred. No.\*62;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 LGPGGRPPPP 11  
   | | | | |  
Db 5 LGPPGSPPPP 14

Search completed: February 5, 2003, 09:47:57  
Job time : 36.4286 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:46:31 ; Search time 15.4286 Seconds  
(without alignments)  
74.771 Million cell updates/sec

Title: US-09-700-329-2

Perfect score: 75

Sequence: 1 ELGPQGRPPPPC 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	66.7	1603	2 S23810	collagen alpha 1(X
2	49	65.3	186	2 A29802	napin precursor (g
3	49	65.3	2241	2 T02857	conserved hypothet
4	48	64.0	211	2 T32976	hypothetical prote
5	48	64.0	327	2 T29031	hypothetical prote
6	48	64.0	373	2 A47234	homeobox protein H
7	47	62.7	463	2 T13425	regulatory protein
8	47	62.7	635	1 WMBEW6	capsid protein - h
9	47	62.7	1546	1 CGHU2E	collagen alpha 2(X
10	47	62.7	3119	2 T49729	HD protein - mouse
11	46	61.3	112	2 D72859	hypothetical prote
12	46	61.3	1010	2 A33509	vinculin - Caenorh
13	45	60.0	315	2 T47181	hypothetical prote
14	45	60.0	363	2 B84565	probable spliceoso
15	45	60.0	375	2 I50389	transcription fact
16	45	60.0	375	2 S35941	homeotic protein H
17	45	60.0	432	2 T23561	hypothetical prote
18	45	60.0	469	2 A24450	collagen alpha 2(V
19	45	60.0	533	2 T10216	hypothetical prote
20	45	60.0	635	2 A57131	collagen alpha 1(X
21	45	60.0	756	2 S74742	exopolysaccharide
22	45	60.0	1019	2 A32856	collagen alpha 1(V
23	45	60.0	1024	1 S18951	collagen alpha 1(X
24	45	60.0	1806	1 CGHU1E	collagen alpha 1(X
25	44	58.7	76	2 C38355	basic proline-rich
26	44	58.7	117	2 D40750	proline-rich prote
27	44	58.7	128	2 D38355	basic proline-rich
28	44	58.7	176	2 A86441	hypothetical prote
29	44	58.7	212	2 B36298	proline-rich prote

30 44 58.7 245 1 W4WL5  
31 44 58.7 251 1 PIHUPF  
32 44 58.7 309 2 S10889  
33 44 58.7 310 1 PIHUSD  
34 44 58.7 392 1 PIHUB6  
35 44 58.7 440 1 S60755  
36 44 58.7 448 1 S29483  
37 44 58.7 452 1 S14332  
38 44 58.7 540 2 S72233  
39 44 58.7 746 2 E83250  
40 44 58.7 867 2 T27136  
41 44 58.7 871 2 T27135  
42 44 58.7 963 2 T19140  
43 44 58.7 1001 2 T28897  
44 44 58.7 1638 2 A42091  
45 44 58.7 1838 1 CGHU1V

E4 protein - human  
salivary proline-r  
proline-rich prote  
salivary proline-r  
salivary proline-r  
rhodopsin - Allote  
rhodopsin [smilar  
rhodopsin - northe  
transcription fact  
still frameshift 3  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
transcription acti  
collagen alpha 1(V

## ALIGNMENTS

## RESULT 1

S23810  
collagen alpha 1(XVI) chain precursor - human  
N:Alternate names: procollagen alpha 1(XVI) chain  
C:Species: Homo sapiens (man)  
C>Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 20-Sep-1999  
C;Accession: S23810; PQ0612; S08012  
R;Pan, T.C.; Zhang, R.Z.; Mattei, M.G.; Timpl, R.; Chu, M.L.  
Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992  
A;Title: Cloning and chromosomal location of human alpha1(XVI) collagen.  
A;Reference number: S23810; MUID:92335339; PMID:1631157  
A;Accession: S23810  
A;Molecule type: mRNA  
A;Residues: 1-1603 <PAN>  
A;Cross-references: EMBL:M92642; NID:g180757; PIDN:AAA58427.1; PTD:g180758  
A;Experimental source: skin fibroblasts  
R;Yamaguchi, N.; Kimura, S.; McBride, O.W.; Hori, H.; Yamada, Y.; Kanamori, T.; Yamak  
J. Biochem. 112, 856-863, 1992  
A;Title: Molecular cloning and partial characterization of a novel collagen chain, al  
A;Reference number: PQ0612; MUID:93203161; PMID:1284248  
A;Accession: PQ0612  
A;Molecule type: mRNA  
A;Residues: 'GGR', 421-536, 'P', 538-1159, 'P', 1161-1162, 'P', 1164, 'P', 1166-1603 <YAM>  
A;Cross-references: GB:S57132; NID:g298641; PIDN:AB25797.1; PTD:g298642  
R;Kimura, S.  
submitted to the EMBL Data Library, April 1989  
A;Description: Partial nucleotide and amino acid sequence of a collagen-like protein  
A;Reference number: S08012  
A;Accession: S08012  
A;Molecule type: mRNA  
A;Residues: 403-419, 'GR', 421-536, 'P', 538-846, 'VM' <KIM>  
A;Cross-references: EMBL:X14963; NID:g29984; PIDN:CAA33085.1; PID:g930048  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating uni  
ed and subsequently O-glycosylated.  
C;Genetics:  
A;Gene: GDB:COL16A1  
A;Cross-references: GDB:I34045; OMIM:120326  
A;Map position: lp34-lp34  
C;Complex: type XVI collagen may be a homotrimer, or a heterotrimer of two alpha 1(XV  
C;Function:  
A;Description: structural component of extracellular fibrous polymer as a minor form  
A;Note: may play a role in forming elastic connections at fibril surfaces  
C;Superfamily: unassigned collagens  
C;Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; hydroxylys  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-1603/Product: collagen alpha 1(XVI) chain #status predicted <MAT>  
F:22-333/Domain: amino-terminal nonhelical #status predicted <NC11>  
F:334-1577/Region: interrupted helical  
F:334-360/Domain: collagenous COL9 #status predicted <COL9>  
F:375-505/Domain: collagenous COL9 #status predicted <COL9>  
F:521-554/Domain: collagenous COL8 #status predicted <COL8>



C:Accession: A47234  
 R:Stadler, H.S.; Padanilam, B.J.; Buetow, K.; Murray, J.C.; Solursh, M.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 11579-11583, 1992  
 A:Title: Identification and genetic mapping of a homeobox gene to the 4p16.1 region of h  
 A:Reference number: A47234; MUID:93087572; PMID:1360670  
 A:Accession: A47234  
 A:Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-373 <STA>  
 A:Experimental source: embryo craniofacial region  
 A:Note: sequence extracted from NCBI backbone (NCBIN:119953, NCBIP:119955)  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F:201-257/Domain: homeobox homology <Hox>

Query Match 64.0%; Score 48; DB 2; Length 373;  
 Best Local Similarity 77.8%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 PQGRPPPPC 12  
 | | | | | | |  
 Db 325 PPSRPPPPC 333

RESULT 7  
 T13425  
 regulatory protein K10, oocyte-specific - fruit fly (*Drosophila melanogaster*)  
 N:Alternate names: protein EG-30B8.5  
 C:Species: *Drosophila melanogaster*  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 15-Oct-1999  
 C:Accession: T13425; A28826  
 R:Murphy, L.; Harris, D.; Barrell, B.  
 submitted to the EMBL Data Library, April 1999  
 A:Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.  
 A:Reference number: 217668  
 A:Accession: T13425  
 A:Status: translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-463 <MUR>  
 A:Cross-references: EMBL:AL009195; NID:el1355203; PID:el202207; PIDN:CAA15702.1  
 R:Prost, E.; Deryckere, F.; Roos, C.; Haenlin, M.; Pantescio, V.; Mohier, E.  
 Genes Dev. 2, 891-900, 1988  
 A:Title: Role of the oocyte nucleus in determination of the dorsoventral polarity of Dro  
 A:Reference number: A28826  
 A:Molecule type: mRNA  
 A:Accession: A28826  
 A:Residues: 1-276, 'HH', 279-281, 'VDHHR', 287-463 <PRO>  
 A:Cross-references: GB:X12836; NID:98148; PID:g295771  
 C:Genetics:  
 A:Gene: FlyBase:fs(1)K10  
 A:Cross-references: FlyBase:FBgn0000810  
 A:Map position: X  
 A:Introns: 432/3  
 C:Keywords: DNA binding

Query Match 62.7%; Score 47; DB 2; Length 463;  
 Best Local Similarity 70.0%; Pred. No. 41;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 LQGRPPPPP 11  
 : | | | | |  
 Db 278 MGPMGGPPPP 287

RESULT 8  
 WHEW6  
 capsid protein - human herpesvirus 1 (strain 17)  
 C:Species: human herpesvirus 1  
 C:Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 16-Jun-2000  
 C:Accession: H30084  
 R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perri  
 J. Gen. Virol. 69, 1531-1574, 1988  
 A:Title: The complete DNA sequence of the long unique region in the genome of herpes sim

A:Reference number: A30083; MUID:88274327; PMID:2839594  
 A:Accession: H30084  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-635 <MCG>  
 A:Cross-references: GB:X14112; NID:g1944536; PIDN:CAA32318.1; PID:g59526; GB:D00317  
 C:Genetics:  
 A:Gene: UL26  
 C:Superfamily: varicella-zoster virus gene 33 protein  
 C:Keywords: capsid protein

Query Match 62.7%; Score 47; DB 1; Length 635;  
 Best Local Similarity 77.8%; Pred. No. 55;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 PQGRPPPPC 12  
 | | | | | | |  
 Db 565 PPGPPPPPC 573

RESULT 9  
 CGHU2E  
 collagen alpha 2(XI) chain precursor - human (fragment)  
 N:Alternate names: procollagen alpha 2(XI) chain  
 N:Contains: proline/arginine-rich protein (PARP)  
 C:Species: *Homo sapiens* (man)  
 C:Date: 07-Jun-1990 #sequence\_revision 03-Oct-1995 #text\_change 22-Jun-1999  
 C:Accession: S34790; A32645  
 R:Zhukova, N.I.; Brewton, R.G.; Mayne, R.  
 FEBS Lett. 326, 25-28, 1993  
 A:Title: Molecular cloning of PARP (proline/arginine-rich protein) from human cartila  
 hain.  
 A:Reference number: S34790; MUID:93314796; PMID:8325374  
 A:Accession: S34790  
 A:Molecule type: mRNA  
 A:Residues: 1-663 <ZHI>  
 A:Cross-references: EMBL:L18987; NID:g306439; PIDN:AAA35498.1; PID:g306440  
 R:Kimura, T.; Cheah, K.S.E.; Chan, S.D.H.; Lui, V.C.H.; Mattei, M.G.; van der Rest, M  
 J. Biol. Chem. 264, 13910-13916, 1989  
 A:Title: The human alpha2(XI) collagen (COL11A2) chain. Molecular cloning of cDNA and  
 A:Reference number: A32645; MUID:89340485; PMID:2760050  
 A:Accession: A32645  
 A:Molecule type: DNA; mRNA  
 A:Residues: 586-1546 <KIM>  
 A:Cross-references: GB:J04974; NID:g180714; PIDN:AAA52034.1; PID:g180715  
 A:Note: parts of this sequence were determined by protein sequencing  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating uni  
 ed and subsequently O-glycosylated.  
 C:Genetics:  
 A:Gene: GDB:COL11A2  
 A:Cross-references: GDB:119788; OMIM:120290  
 A:Map position: 6p21.3-6p21.3  
 A:Introns: 1302/3; 1320/3; 1332/3; 1350/3; 1440/1; 1477/3  
 A:Note: the list of introns is incomplete  
 C:Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains (see PIR:  
 3(XI) chain (see PIR:CGHU6C), initially linked by disulfide bonds among their carbox  
 rmed with desmosine cross-links made from lysine and allysine residues  
 C:Function:  
 A:Description: structural component of extracellular fibrous polymer associated with  
 A:Note: may play a role in controlling the lateral growth of collagen II fibrils  
 C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homolo  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxypr  
 F:1-254/Domain: non-collagenous (fragment) #status predicted <NC3>  
 F:1-187/Product: proline/arginine-rich PARP protein (fragment) #status predicted <PAR  
 F:255-305/Domain: collagenous, triple helix #status predicted <COL2>  
 F:306-342/Domain: non-collagenous #status predicted <NC2>  
 F:343-1356/Region: helical  
 F:429-431/Region: cell attachment (R-G-D) motif  
 F:447-449/Region: cell attachment (R-G-D) motif  
 F:1257-1259/Region: cell attachment (R-G-D) motif  
 F:1357-1380/Region: carboxyl-terminal nonhelical telopeptide  
 F:1381-1546/Domain: carboxyl-terminal propeptide (fragment) #status predicted <CTP>  
 F:1403-1546/Domain: fibrillar collagen carboxyl-terminal homology (fragment) #status

F:109-163,1511-1545/Disulfide bonds: #status predicted  
 F:319/Modified site: allysine (lys) #status predicted  
 F:426,1266/Modified site: 5-hydroxylysine (lys) #status predicted  
 F:426,1266/Binding site: carbohydrate (lys) (covalent) #status predicted  
 F:927,933,1008,1017,1035,1038,1290,1296,1305,1317,1320/Modified site: 4-hydroxyproline (pro) #status atypical  
 F:929/Modified site: 4-hydroxyproline (pro) #status atypical  
 F:942,1023,1299/Modified site: 5-hydroxylysine (lys) #status experimental  
 F:942,1023,1299/Binding site: carbohydrate (lys) (covalent) #status experimental  
 F:1427,1433,1450,1459/Disulfide bonds: interchain #status predicted  
 F:1460/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 62.7%; Score 47; DB 1; Length 1546;  
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 LGPQGRPPPP 11  
 DB 495 LGPQGEPP 504

RESULT 10  
 I49729  
 HD protein - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 05-Nov-1999  
 C:Accession: I49729; I49730  
 R:Lin, B.; Nasir, J.; MacDonald, H.; Hutchinson, G.; Graham, R.; Rommens, J.M.; Hayden, H.M. Mol. Genet. 3, 85-92, 1994  
 A:Title: Sequence of the murine huntington disease gene: evidence for conservation and a  
 A:Reference number: I49729; MUID:94214482; PMID:8162057  
 A:Accession: I49729  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3119 <RES>  
 A:Cross-references: GB:L23312; NID:g438804; PIDN:AAA37799.1; PID:g438805  
 A:Accession: I49730  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1521,2002-3119 <RE2>  
 A:Cross-references: GB:L23313; NID:g438806; PIDN:AAA37800.1; PID:g438807  
 C:Keywords: alternative splicing

Query Match 62.7%; Score 47; DB 2; Length 3119;  
 Best Local Similarity 87.5%; Pred. No. 2.4e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 POGRRPPPP 11  
 DB 45 POGQPPPP 52

RESULT 11  
 D72654  
 hypothetical protein APE0664 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C:Accession: D72654  
 R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
 DNA Res. 6, 83-101, 1999  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
 A:Reference number: A72450; MUID:99310339; PMID:10382966  
 A:Accession: D72654  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-112 <KAW>  
 A:Cross-references: DBJ:AP000060; NID:g5104188; PIDN:BAA79636.1; PID:d1043422; PID:g510  
 A:Experimental source: strain K1  
 C:Genetics:  
 A:Gene: APE0664

Query Match 61.3%; Score 46; DB 2; Length 112;  
 Best Local Similarity 77.8%; Pred. No. 15;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 POGRRPPPP 12  
 DB 71 POGRRPPSP 79

RESULT 12  
 A33509  
 vinculin - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 22-Nov-1989 #sequence\_revision 22-Nov-1989 #text\_change 01-Mar-2002  
 C:Accession: A33509  
 R:Barstead, R.J.; Waterston, R.H.  
 J. Biol. Chem. 264, 10177-10185, 1989  
 A:Title: The basal component of the nematode dense-body is vinculin.  
 A:Reference number: A33509; MUID:89255506; PMID:2498337  
 A:Accession: A33509  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1010 <BAR>  
 A:Cross-references: GB:J04804  
 C:Superfamily: meta-vinculin; vinculin amino-terminal homology; vinculin carboxyl-ter  
 C:Keywords: actin binding; cytoskeleton  
 F:4-257/Domain: vinculin amino-terminal homology <VINC>  
 F:477-1007/Domain: vinculin carboxyl-terminal homology <VINC>

Query Match 61.3%; Score 46; DB 2; Length 1010;  
 Best Local Similarity 72.7%; Pred. No. 1.1e+02;  
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ELGPQGRPPPP 11  
 DB 796 ELSPRRPPPP 806

RESULT 13  
 T47181  
 hypothetical protein DKFZp434F0616.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 08-Dec-2000  
 C:Accession: T47181  
 R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Weil, B.; Wiemann, S.  
 Submitted to the Protein Sequence Database, March 2000  
 A:Reference number: Z24378  
 A:Accession: T47181  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-315 <AAA>  
 A:Cross-references: EMBL:AL162003  
 A:Experimental source: adult testis; clone DKFZp434F0616  
 C:Genetics:  
 A:Note: DKFZp434F0616.1  
 C:Superfamily: POZ domain homology

Query Match 60.0%; Score 45; DB 2; Length 315;  
 Best Local Similarity 70.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GPOGRPPPP 12  
 DB 86 GPYGEPPYC 95

RESULT 14  
 B84565  
 probable spliceosome associated protein [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cross)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 17-May-2002  
 C:Accession: B84565  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y  
 M.; Koo, H.; Morfat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: AB4420; MUID:20083487; PMID:10617197  
A:Accession: B84565

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-363 <STO>

A:Cross-references: GB:AB002093; NID:g4218014; PIDN:AAD12222.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g18510

A:Map position: 2

C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein

Query Match 60.0%; Score 45; DB 2; Length 363;  
Best Local Similarity 87.5%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 PQGRPPPP 11

DB 291 PQMPPPPP 298

#### RESULT 15

I50589

transcription factor - chicken

C:Species: Gallus gallus (chicken)

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 19-Jul-2002

C:Accession: I50589

R:Prince, V.; Lumsden, A.

Development 120, 911-923, 1994

A:Title: Hoxa-2 expression in normal and transposed rhombomeres: independent regulation

A:Reference number: I50589; MUID:95324377; PMID:7600967

A:Accession: I50589

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-375 <PRI>

A:Cross-references: EMBL:X74323; NID:g415799; PIDN:CAA52370.1; PID:g415800

C:Genetics:

A:Gene: Hoxa-2

C:Superfamily: homeotic protein Hox B2; homeobox homology

C:keywords: DNA binding; homeobox; nucleus; transcription regulation

F:141-197/Domain: homeobox homology <HOX>

Query Match 60.0%; Score 45; DB 2; Length 375;  
Best Local Similarity 77.8%; Pred. No. 62;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GPQGRPPPP 11

DB 85 GPAGAPPPP 93

Search completed: February 5, 2003, 09:49:49  
Job time : 17.4286 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:41:15 ; Search time 8.57143 Seconds  
(without alignments)  
58.067 Million cell updates/sec

Title: US-09-700-329-2  
Perfect score: 75  
Sequence: 1 ELGQGRPPPC 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	51	68.0	351	1 KLF2_RAT	Q9t58 rattus norv
2	51	68.0	354	1 KLF2_MOUSE	Q60843 mus musculus
3	50	66.7	355	1 KLF2_HUMAN	Q9Y5W3 homo sapien
4	50	66.7	1603	1 CALF_HUMAN	Q07092 homo sapien
5	49	65.3	186	1 2SSE_BRANA	P09893 brassica na
6	47	62.7	463	1 K10_DROME	P13468 drosophila
7	47	62.7	635	1 VP40_HSV11	P10210 herpes simp
8	47	62.7	1650	1 CA2B_MOUSE	Q64739 mus musculus
9	47	62.7	1736	1 CA2B_HUMAN	P13942 homo sapien
10	47	62.7	3110	1 HD_RAT	P51111 rattus norv
11	47	62.7	3119	1 HD_MOUSE	P42859 mus musculus
12	46	61.3	482	1 BTB1_HUMAN	Q9H0C5 homo sapien
13	46	61.3	1010	1 VINC_CAEEL	P19826 caenorhabdi
14	45	60.0	375	1 HX2_CHICK	Q08727 gallus gall
15	45	60.0	615	1 HIC2_HUMAN	Q961B3 homo sapien
16	45	60.0	635	1 CA28_HUMAN	P25067 homo sapien
17	45	60.0	911	1 CALB_BOVIN	Q28083 bos taurus
18	45	60.0	1019	1 CAL6_CHICK	P20785 gallus gall
19	45	60.0	1804	1 CALB_MOUSE	Q61245 mus musculus
20	45	60.0	1806	1 CALB_HUMAN	P12107 homo sapien
21	44	58.7	61	1 PRPE_HUMAN	P02811 homo sapien
22	44	58.7	96	1 PRP5_HUMAN	P04281 homo sapien
23	44	58.7	234	1 PRPM_HUMAN	P10161 homo sapien
24	44	58.7	245	1 VE4_HPV05	P06924 human papil
25	44	58.7	247	1 PRB4_HUMAN	P10163 homo sapien
26	44	58.7	251	1 PRP2_HUMAN	P02812 homo sapien
27	44	58.7	276	1 PRPL_HUMAN	P10162 homo sapien
28	44	58.7	331	1 PRP1_HUMAN	P04280 homo sapien
29	44	58.7	439	1 OPSD_LOLSU	Q17094 loligo subu
30	44	58.7	447	1 OPSD_TODPA	P13356 todarodes p
31	44	58.7	452	1 OPSD_LOLFO	P24603 loligo forb
32	44	58.7	464	1 OPSD_SEPOF	O16005 sepia offic
33	44	58.7	540	1 TBX6_MOUSE	P70327 mus musculus

34	44	58.7	963	1 YQ36_CAEEL	Q09457 caenorhabdi
35	44	58.7	1638	1 BRM_DROME	P25439 drosophila
36	44	58.7	1838	1 CAL5_HUMAN	P20908 homo sapien
37	43.5	58.0	676	1 TIM_DROHY	O44431 drosophila
38	43.5	58.0	1343	1 TIM_DROVI	Q17482 drosophila
39	43.5	58.0	1421	1 TIM_DROME	P49021 drosophila
40	43	57.3	345	1 CC14_CAEEL	P18834 caenorhabdi
41	43	57.3	860	1 ELS_MOUSE	P54320 mus musculus
42	43	57.3	864	1 ELS_RAT	Q99372 rattus norv
43	43	57.3	1022	1 CA26_CHICK	P15988 gallus gall
44	43	57.3	1144	1 RIR1_HSV23	P09853 herpes simp
45	42	56.0	112	1 VPX_HV2N2	P05915 human immun

ALIGNMENTS

RESULT 1  
KLF2\_RAT  
ID KLF2\_RAT STANDARD; PRT; 351 AA.  
AC Q9ET58;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Kruppel-like factor 2 (lung kruppel-like factor).  
GN KLF2 OR LKLF.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BN/SsNHsd;  
RA Haag F., Bartels K., Rothenburg S., Stahmer I., Thiele H.-G.,  
RA Koch-Nolte F.;  
RT "The gene for the transcription factor LKLF is developmentally  
expressed in rat T cells and is not defective in lymphopenic  
diabetes-prone BB rats.";  
RT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RL Submitted (AUG-1999) to the CACCC BOX IN THE BETA-GLOBIN GENE PROMOTER  
CC -1- FUNCTION: BINDS TO THE CACCC BOX IN THE BETA-GLOBIN GENE PROMOTER  
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
FINGER PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; AF181251; AAG02141.1; -.  
CC HSSP; P08047; 1SP2.  
DR InterPro: IPR000822; Znf\_C2H2.  
DR Pfam; PF00096; zf\_C2H2; 3.  
DR ProDom; PD000003; Znf\_C2H2; 2.  
DR SMART; SM00355; Znf\_C2H2; 3.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 3.  
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;  
KW DNA-binding; Nuclear protein; Repeat.  
FT DOMAIN 60 70  
FT DOMAIN 165 168  
FT DOMAIN 224 227  
FT DOMAIN 268 350  
FT ZN\_FING 268 292  
FT ZN\_FING 298 322  
FT ZN\_FING 328 350  
SQ SEQUENCE 351 AA; 37313 MW; 2E4EB6B0577A53A4 CRC64;

Query Match 68.0%; Score 51; DB 1; Length 351;

```

Best Local Similarity 88.9%; Pred. No. 5.8;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GPOGRPPPP 11
  || |||||
Db 160 GPAGRPPPP 168

RESULT 2
KLF2_MOUSE STANDARD; PRT; 354 AA.
ID KLF2_MOUSE
AC Q60843;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kruppel-like factor 2 (Lung kruppel-like factor).
GN KLF2 OR IKLF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=96025976; PubMed=7565748;
RA Anderson K.P., Kern C.B., Crable S.C., Lingrel J.B.;
RT "Isolation of a gene encoding a functional zinc finger protein
RT homologous to erythroid Kruppel-like factor: identification of a new
RT multigene family.";
RL Mol. Cell. Biol. 15:5957-5965(1995).
CC -!- FUNCTION: BINDS TO THE CACCC BOX IN THE BETA-GLOBIN GENE PROMOTER
CC AND ACTIVATES TRANSCRIPTION.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- TISSUE SPECIFICITY: PREDOMINANT EXPRESSION IN THE LUNGS AND
CC SPLEEN.
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U25096; AAA86728.1; -.
CC HSSP; P08047; ISP2.
CC MGD; MGI:1342772; Klf2.
CC InterPro; IPR000822; Znf_C2H2.
CC Pfam; PF00059; zf-C2H2_3.
CC PRINTS; PR00048; ZINCFINGER.
CC ProDom; PD000003; Znf_C2H2; 2.
CC SMART; SM00355; Znf_C2H2; 3.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
CC Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT DOMAIN 60 70 POLY-PRO.
FT DOMAIN 165 169 POLY-PRO.
FT DOMAIN 224 230 POLY-ALA.
FT DOMAIN 271 353 ZINC FINGERS.
FT ZN_FING 271 295 ZINC FINGERS.
FT ZN_FING 301 325 C2H2-TYPE.
FT ZN_FING 331 353 C2H2-TYPE.
FT ZN_FING 331 353 C2H2-TYPE.
SQ SEQUENCE 354 AA; 37700 MW; C4A9DD018AC5BAF7 CRC64;

Query Match 68.0%; Score 51; DB 1; Length 354;
Best Local Similarity 88.9%; Pred. No. 5.9;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GPOGRPPPP 11
  || |||||
Db 160 GPAGRPPPP 168

RESULT 3
KLF2_HUMAN STANDARD; PRT; 355 AA.
ID KLF2_HUMAN
AC Q9Y5W3; Q9UKR6; Q9UJS5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Kruppel-like factor 2 (Lung kruppel-like factor).
GN KLF2 OR IKLF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=99231781; PubMed=10217429;
RA Kisselev L.L.;
RT "Structure of the human CpG-island-containing lung Kruppel-like factor
RT (IKLF) gene and its location in chromosome 19p13.11-13 locus.";
RL FEBS Lett. 448:149-152(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=99389728; PubMed=10458913;
RA Wani M.A., Konkright M.D., Jeffries S., Hughes M.J., Lingrel J.B.;
RT "cDNA isolation, genomic structure, regulation, and chromosomal
RT localization of human lung kruppel-like factor.";
RL Genomics 60:78-86(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Lee H.J., Kim M.K., Kim Y.H., Seo J.M., Lee H.M., Chung H.J.,
RA Sohn M.Y., Hwang S.Y., Im S.U., Jung E.J., Kim J.C.;
RT "A catalogue of genes in the human dermal papilla cells as identified
RT by expressed sequence tags.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS TO THE CACCC BOX IN THE BETA-GLOBIN GENE PROMOTER
CC AND ACTIVATES TRANSCRIPTION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF123344; AAD25076.1; -.
CC EMBL; AF134053; AAD55891.1; -.
CC EMBL; AF205849; AAF13295.1; -.
CC HSSP; P08047; ISP2.
CC TRANSFAC; T04958; -.
CC Genew; HGNC:6347; KLF2.
CC MIM; 602016; -.
CC InterPro; IPR000822; Znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 3.
CC ProDom; PD000003; Znf_C2H2; 2.
CC SMART; SM00355; Znf_C2H2; 3.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
CC Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT DOMAIN 62 71 POLY-PRO.
FT DOMAIN 130 135 POLY-GLY.
FT DOMAIN 167 171 POLY-PRO.

```

```
FT DOMAIN 225 231 POLY-ALA.
FT DOMAIN 272 354 ZINC FINGERS.
FT ZN_FING 272 296 C2H2-TYPE.
FT ZN_FING 302 326 C2H2-TYPE.
FT ZN_FING 332 354 C2H2-TYPE.
FT CONFLICT 43 43 S -> N (IN REF. 2).
FT CONFLICT 104 104 L -> P (IN REF. 1).
FT CONFLICT 175 175 P -> S (IN REF. 2).
FT CONFLICT 184 184 L -> M (IN REF. 2).
SQ SEQUENCE 355 AA; D5849C831D676AE1 CRC64;

Query Match 66.7%; Score 50; DB 1; Length 355;
Best Local Similarity 88.9%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GPQGRPPPP 11
Db 162 GPQGRPPPP 170

RESULT 4
CALF_HUMAN
ID CALF_HUMAN STANDARD; PRT; 1603 AA.
AC Q07092.
DC 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XVI) chain precursor.
GN COL16A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333339; PubMed=1631157;
RA Pan T.C., Zhang R.Z., Mattei M.-G., Timpl R., Chu M.-L.;
RT "Cloning and chromosomal location of human alpha 1(XVI) collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6565-6569(1992).
RN [2]
RP SEQUENCE OF 418-1603 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93203161; PubMed=1284248;
RA Yanaguchi N., Kimura S., McBride O.W., Hori H., Yamada Y.,
RA Kanamori T., Yamakoshi H., Nagai Y.;
RT "Molecular cloning and partial characterization of a novel collagen
chain, alpha 1(XVI), consisting of repetitive collagenous domains and
cysteine-containing non-collagenous segments.";
RL J. Biochem. 112:856-863(1992).
CC -!- FUNCTION: THE NUMEROUS INTERRUPTIONS IN THE TRIPLE HELIX MAY MAKE
CC -!- TISSUE SPECIFICITY: IN THE PLACENTA, WHERE IT IS FOUND IN THE
CC ANNION, A MEMBRANOUS TISSUE LINING THE AMNIOIC CAVITY. WITHIN THE
CC ANNION, IT IS FOUND IN AN ACCELLULAR, RELATIVELY DENSE LAYER OF A
CC COMPLEX NETWORK OF RETICULAR FIBERS. ALSO LOCATED TO A FIBROBLAST
CC LAYER BENEATH THIS DENSE LAYER. EXISTS IN TISSUES IN ASSOCIATION
CC WITH OTHER TYPES OF COLLAGEN.
CC -!- DEVELOPMENTAL STAGE: TRANSIENTLY ELEVATED EXPRESSION DURING
CC GESTATION, AND DECREASE AT TERM
CC -!- DOMAIN: THIS SEQUENCE DEFINES EIGHTEEN DIFFERENT DOMAINS, NINE
CC TRIPLE-HELICAL DOMAINS (COL9 TO COL1) AND TEN NONTRIPLE-HELICAL
CC DOMAINS (NC10 TO NC1).
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH ,
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M92642; AAA58427.1; -.
DR EMBL; S57132; AAB25797.1; -.
DR PIR; S23810; S23810.
DR Genew; HGNC:2193; COL16A1.
DR MIM; 120326; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00210; TSPN; 1.
KW Extracellular matrix; Connective tissue; Collagen; Hydroxylation;
KW Repeat; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1603 COLLAGEN ALPHA 1(XVI) CHAIN.
FT DOMAIN 22 374 NONHELICAL REGION 10 (NC10).
FT DOMAIN 375 505 TRIPLE-HELICAL REGION 9 (COL9)
WITH 3 IMPERFECTIONS.
FT DOMAIN 506 520 NONHELICAL REGION 9 (NC9).
FT DOMAIN 521 554 TRIPLE-HELICAL REGION 8 (COL8)
WITH 1 IMPERFECTION.
FT DOMAIN 555 571 NONHELICAL REGION 8 (NC8).
FT DOMAIN 572 630 TRIPLE-HELICAL REGION 7 (COL7)
WITH 1 IMPERFECTION.
FT DOMAIN 631 651 NONHELICAL REGION 7 (NC7).
FT DOMAIN 652 722 TRIPLE-HELICAL REGION 6 (COL6)
WITH 1 IMPERFECTION.
FT DOMAIN 723 737 NONHELICAL REGION 6 (NC6).
FT DOMAIN 738 875 TRIPLE-HELICAL REGION 5 (COL5)
WITH 3 IMPERFECTIONS.
FT DOMAIN 876 886 NONHELICAL REGION 5 (NC5).
FT DOMAIN 887 938 TRIPLE-HELICAL REGION 4 (COL4)
WITH 2 IMPERFECTIONS.
FT DOMAIN 939 972 NONHELICAL REGION 4 (NC4).
FT DOMAIN 973 987 TRIPLE-HELICAL REGION 3 (COL3).
FT DOMAIN 988 1010 NONHELICAL REGION 3 (NC3).
FT DOMAIN 1011 1432 TRIPLE-HELICAL REGION 2 (COL2)
WITH 2 IMPERFECTIONS.
FT DOMAIN 1433 1471 NONHELICAL REGION 2 (NC2).
FT DOMAIN 1472 1577 TRIPLE-HELICAL REGION 1 (COL1)
WITH 2 IMPERFECTIONS.
FT DOMAIN 1578 1603 NONHELICAL REGION 1 (NC1).
FT CONFLICT 418 420 RDA -> GGR (IN REF. 2).
FT CONFLICT 537 537 R -> P (IN REF. 2).
FT CONFLICT 1160 1160 T -> P (IN REF. 2).
FT CONFLICT 1163 1163 T -> P (IN REF. 2).
FT CONFLICT 1165 1165 S -> P (IN REF. 2).
SQ SEQUENCE 1603 AA; E27D9A1D4E598A37 CRC64;

Query Match 66.7%; Score 50; DB 1; Length 1603;
Best Local Similarity 80.0%; Pred. No. 31;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LGPQGRPPPP 11
Db 1280 MGPQGRPPPP 1289

RESULT 5
2SSE_BRANA
ID 2SSE_BRANA STANDARD; PRT; 186 AA.
AC P09893;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Napin embryo specific precursor (1.7S seed storage protein).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=87308225; PubMed=3040733;
RA Scofield S.R., Crouch M.L.;
RT "Nucleotide sequence of a member of the napin storage protein family
from Brassica napus.";
RL J. Biol. Chem. 262:12202-12208(1987).
CC -!- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE
CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
CC ITS MATURATION.
CC -!- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -!- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
CC -!- DEVELOPMENTAL STAGE: EMBRYO.
CC -!- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02782; AAA33007.1; -.
DR PIR; A29802; A29802.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR000617; Napin.
DR InterPro; IPR001768; Try/amyL_1bhbtr.
DR Pfam; PF00234; tryp_alpha_aml1; 1.
DR PRINTS; PR00496; NAPIN.
DR ProDom; PD002498; Napin; 1.
DR SMART; SM00499; AAI; 1.
DR Seed storage protein; Signal; Multigene family.
KW SIGNAL 1 21
FT PROPEP 22 38 SMALL CHAIN.
FT CHAIN 39 76
FT PROPEP 77 97
FT CHAIN 98 186 LARGE CHAIN.
FT CHAIN 186 AA; 21013 MW; 9CAE63D84B160AB3 CRC64;
SQ SEQUENCE 186 AA; 21013 MW; 9CAE63D84B160AB3 CRC64;

Query Match 65.3%; Score 49; DB 1; Length 186;
Best Local Similarity 88.9%; Pred. No. 5.9;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GPQGRPPPP 11
DB 97 GPQGRPPPP 105
|||||

RESULT 6
K10_DROME STANDARD; PRT; 463 AA.
AC P13468; O46075; O9W505;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-binding protein K10 (Female sterile protein K10).
GN FS(1)K10 OR EG:3088.5 OR CG3218.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Prost E., Deryckere F., Roos C., Haenlin M., Pantescio V.,
RA Mohler V.;
RT "Role of the oocyte nucleus in determination of the dorsoventral
RT polarity of Drosophila as revealed by molecular analysis of the K10
RT gene.";
RL Genes Dev. 2:891-900(1988).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=Oregon-R;
RX MEDLINE=20196011; PubMed=10731137;
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
RA Minana B., Glouf S., Lelaure V., Mottier S., Galibert F., Borkova D.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., Bolshakov S.,
RA Modolell J., Peter A., Schoettler P., Werner M., Mourikioti F.,
RA Belinier D.M., Dowe G., Schaefer U., Jaeckle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
RA McMillan P.J., Sallies C., Tait E.A., Valenti P., Saunders R.D.C.,
RA Glover D.M.;
RT "From sequence to chromosome: the tip of the X chromosome of D.
RT melanogaster.";
RL Science 287:2220-2222(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Posler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: MAY HAVE A REGULATORY FUNCTION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X12836; CAA31321.1; -.
DR EMBL; AL009195; CAA15702.1; -.
DR EMBL; AE003423; AAF45758.1; -.
DR PIR; A28826; A28826.
DR FlyBase; FBgn0000810; fs(1)K10.

```

KW DNA-binding: Nuclear protein; Repeat.  
FT DOMAIN 87 142 7 X APPROXIMATE TANDEM REPEATS.  
FT REPEAT 87 94 1.  
FT REPEAT 95 102 2.  
FT REPEAT 103 110 3.  
FT REPEAT 111 118 4.  
FT REPEAT 119 126 5.  
FT REPEAT 127 134 6.  
FT REPEAT 135 142 7.  
FT DOMAIN 284 290 POLY-PRO.  
FT DNA\_BIND 397 416 H-T-H MOTIF (POTENTIAL).  
FT CONFLICT 277 278 PM -> HH (IN REF. 1).  
FT CONFLICT 282 286 GGPPP -> VDHHR (IN REF. 1).  
SQ SEQUENCE 463 AA; 51267 MW; D03C097192D1FDD0 CRC64;  
  
Query Match 62.7%; Score 47; DB 1; Length 463;  
Best Local Similarity 70.0%; Pred. No. 24;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 LGPQGRPPPP 11 :||| ||||  
DB 278 MGPMGPPPP 287  
  
RESULT 7  
ID VP40\_HSV11 STANDARD; PRT; 635 AA.  
AC P10210;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Capsid protein P40 (Viral structural protein UL26) [Contains: Capsid  
DE protein VP24 (Assemblin) (Protease) (EC 3.4.21.97); Capsid protein  
DE VP22A].  
GN UL26.  
OS Herpes simplex virus (type 1 / strain 17).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Simplexvirus.  
OX NCBI\_TaxID=10299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88274327; PubMed=2839594;  
RA McGeoch D.J., Dairymple M.A., Davison A.J., Dolan A., Frame M.C.,  
RA McNab D., Perry L.J., Scott J.E., Taylor P.;  
RT "The complete DNA sequence of the long unique region in the genome of  
RT herpes simplex virus type 1".  
RL J. Gen. Virol. 69:1531-1574(1988).  
RN [2]  
RP SEQUENCE OF 11-29; 77-91 AND 223-241.  
RX MEDLINE=93019027; PubMed=1328483;  
RA Davison M.D., Rixon F.J., Davison A.J.;  
RT "Identification of genes encoding two capsid proteins (VP24 and VP26)  
RT of herpes simplex virus type 1".  
RL J. Gen. Virol. 73:2709-2713(1992).  
CC -!- FUNCTION: VP22A IS A COMPONENT OF THE CAPSID CORE INVOLVED IN  
CC PROCESSING AND PACKAGING OF PROGENY DNA. VP24 IS A PROTEASE  
CC WHICH CAN PROTEOLYTICALLY CLEAVE ITSELF AND VP22A AT THE C-  
CC TERMINUS.  
CC -!- CATALYTIC ACTIVITY: Preferentially cleaves at Ala-I-Ser or Ala-I-  
CC Ala.  
CC -!- PTM: VP22A IS PHOSPHORYLATED.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S21.  
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
EMBL; D10879; BAA01672.1; -.

DR EMBL: X14112; CAA32318.1; -.  
DR PIR; H30084; WHBEM6.  
DR HSP; P16753; IWPO.  
DR MEROPS; S21.001; -.  
DR InterPro; IPR001847; Assemblin.  
DR Pfam; PF00716; Peptidase\_S21.1.  
DR PRINTS; PR00236; HSCAPSIDP40.  
KW Coat protein; Hydrolase; Serine protease; Phosphorylation.  
FT CHAIN 1 635 GENE UL26 PROTEIN.  
FT CHAIN 253 635 GENE UL26.5 PROTEIN.  
FT CHAIN 1 247 CAPSID PROTEIN VP24 (PROTEASE).  
FT CHAIN 248 ?610 CAPSID PROTEIN VP22A.  
FT PROPEP ?611 635 C-TERMINAL PEPTIDE.  
FT SITE 247 248 CLEAVAGE (BY THE PROTEASE) (PROBABLE).  
FT SITE 610 611 CLEAVAGE (BY THE PROTEASE) (PROBABLE).  
FT ACT\_SITE 61 61 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 129 129 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 148 148 CHARGE RELAY SYSTEM (BY SIMILARITY).  
SQ SEQUENCE 635 AA; 66470 MW; F3B3C7D42F3D062D CRC64;  
  
Query Match 62.7%; Score 47; DB 1; Length 635;  
Best Local Similarity 77.8%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 4 PGRPPPPPP 12 :||| ||||  
DB 565 PGPPPPPPPP 573  
  
RESULT 8  
ID CA2B\_MOUSE STANDARD; PRT; 1650 AA.  
AC Q64739; Q921W0;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Collagen alpha 2(XI) chain precursor.  
GN COL11A2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SVJ;  
RA Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L.,  
RA Hall J., Lasky S., Hood L.;  
RT "Sequence of the mouse major histocompatibility locus class II  
RT region".  
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE OF 1-1592 FROM N.A.  
RC STRAIN=FVB/N, and 129/SV; TISSUE=Cartilage;  
RX MEDLINE=971335795; PubMed=8981332;  
RA Vandenberg P., Vuoristo M., Ala-Kokko L., Prockop D.J.;  
RT "The mouse coll1a2 gene. Some transcripts from the adjacent rxr-beta  
RT gene extend into the coll1a2 gene".  
RL Matrix Biol. 15:359-367(1996).  
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY  
CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS (BY SIMILARITY).  
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),  
CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL  
CC MODIFICATION OF ALPHA 1(XI). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD  
CC OF ALPHA 3(XI)-1(XI) (BY SIMILARITY).  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS ARE PRODUCED BY  
CC ALTERNATIVE SPLICING.  
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC EMBL; AF100956; AAC69905.1; -.
CC EMBL; U16789; AAA67751.1; -.
CC EMBL; U16790; AAA67752.1; -.
CC MGI; 88447; Coll1a2
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001791; Laminin-G.
DR InterPro; IPR001230; Prenyl_site.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Alternative splicing; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1414
FT PROPEP 1415 1650
FT DOMAIN 23 225
FT DOMAIN 487 1414
FT DOMAIN 1415 1650
FT CONFLICT 450 450
FT CONFLICT 618 619
FT CONFLICT 711 711
FT CONFLICT 757 759
FT CONFLICT 768 768
FT CONFLICT 790 790
FT CONFLICT 803 803
FT CONFLICT 836 836
FT CONFLICT 919 919
FT CONFLICT 1167 1167
FT CONFLICT 1300 1300
FT CONFLICT 1436 1436
SQ SEQUENCE 1650 AA; 162072 MW; 8ECDB8702E71E2DA CRC64;
Query Match 62.7%; Score 47; DB 1; Length 1650;
Best Local Similarity 80.0%; Pred. No. 77;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 LGPQGRPPPP 11
DB 553 LGPQGP GPP 562
RESULT 9
CA2B_HUMAN STANDARD; PRT; 1736 AA.
AC P13942; Q13273; Q13271; Q13272; Q07751; Q99866; Q9UIP9;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 2(XI) chain precursor.
GN COL11A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96032717; PubMed=7559422;
RA Vuorio M.M., Pihlajamaa T., Vandenbergh P., Prockop D.J.,
RA Ala-Kokko L.;
RT "The human COL11A2 gene structure indicates that the gene has not
RT evolved with the genes for the major fibrillar collagens.";
```

```

RL J. Biol. Chem. 270:22873-22881(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Tubby B.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 59-807 FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=93314796; PubMed=8325374;
RA Zhidkova N.I., Brewton R.G., Mayne R.;
RT "Molecular cloning of PARP (proline/arginine-rich protein) from human
RT cartilage and subsequent demonstration that PARP is a fragment of the
RT NH2-terminal domain of the collagen alpha 2(XI) chain.";
RL FEBS Lett. 326:25-28(1993).
RN [4]
RP SEQUENCE OF 730-1690 FROM N.A.
RX MEDLINE=89340485; PubMed=2760050;
RA Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mattei M.-G.,
RA van der Rest M., Ono K., Solomon E., Ninomiya Y., Olsen B.R.;
RT "The human alpha 2(XI) collagen (COL11A2) chain. Molecular cloning of
RT cDNA and genomic DNA reveals characteristics of a fibrillar collagen
RT with differences in genomic organization.";
RL J. Biol. Chem. 264:13910-13916(1989).
RN [5]
RP SEQUENCE OF 1-537 FROM N.A.
RX MEDLINE=96435918; PubMed=8838804;
RA Lui V.C., Ng L.J., Sat E.W., Cheah K.S.;
RT "The human alpha 2(XI) collagen gene (COL11A2): completion of coding
RT information, identification of the promoter sequence, and precise
RT localization within the major histocompatibility complex reveal
RT overlap with the KES gene.";
RL Genomics 32:401-412(1996).
RN [6]
RP ALTERNATIVE SPLICING.
RX MEDLINE=95238468; PubMed=7721876;
RA Zhidkova N.I., Justice S.K., Mayne R.;
RT "Alternative mRNA processing occurs in the variable region of the
RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";
RL J. Biol. Chem. 270:9486-9493(1995).
RN [7]
RP DISEASE.
RX PubMed=10677296;
RA Melkonian M., Brunner H.G., Manouvrier S., Hennekam R.,
RA Superfili-Furga A., Kaeferlaenen H., Pauli R.M., van Essen T.,
RA Warman M.L., Bonaventure J., Miny P., Ala-Kokko L.;
RT "Autosomal recessive disorder of osseous dysplasia is
RT associated with loss-of-function mutations in the COL11A2 gene.";
RL Am. J. Hum. Genet. 66:368-377(2000).
RN [8]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [9]
RP VARIANT OSMD ARG-661.
RX MEDLINE=95163096; PubMed=7859284;
RA Vakkula M., Mariman E.C.M., Lui V.C.H., Zhidkova N.I., Tiller G.E.,
RA Goldring M.B., van Beersum S.E.C., de Waal Malefijt M.C.,
RA van den Hoogen F.H.J., Rogers H.-H., Mayne R., Cheah K.S.E.,
RA Olsen B.R., Warman M.L., Brunner H.G.;
RT "Autosomal dominant and recessive osteochondrodysplasias associated
RT with the COL11A2 locus.";
RL Cell 80:431-437(1995).
RN [10]
RP VARIANTS GLY-593; LYS-824; LEU-879; THR-1316 AND GLN-1600.
RX PubMed=9585596;
RA Koga H., Sakou T., Taketomi E., Hayashi K., Numasawa T., Harata S.,
RA Yone K., Matsunaga S., Otterud B., Inoue I., Leppert M.;
RT "Genetic mapping of ossification of the posterior longitudinal
RT ligament of the spine.";
```

RL Am. J. Hum. Genet. 62:1460-1467(1998).  
RN [11]  
RP VARIANT WZS GLU-955.  
RX PubMed-9805126; Prockop D.J., Faber J., Winterpacht A., Zabel B.,  
RA Pihlajama T., Brunner H.G., Creemers C.W.R.J., Takano M., Li S.-W.,  
RA Giedion A., Wiesbauer P., Spranger J., Ala-Kokko L.;  
RT "Heterozygous glycine substitution in the COL11A2 gene in the original  
RT patient with the Weissenbacher-Zweymueller syndrome demonstrates its  
RT identity with heterozygous OSMED (nonocular Stickler syndrome).";  
RL Am. J. Med. Genet. 80:115-120(1998).  
RN [12]  
RP VARIANT STL3 940-GLY--PRO-948 DEL.  
RX PubMed-9506662;  
RA Sirko-Osada D.A., Murray M.A., Scott J.A., Lavery M.A., Warman M.L.,  
RA Robin N.H.;  
RT "Stickler syndrome without eye involvement is caused by mutations in  
RT COL11A2, the gene encoding the alpha-2(XI) chain of type XI  
RT collagen.";  
RL J. Pediatr. 132:368-371(1998).  
RN [13]  
RP VARIANTS DFNA13 GLU-808 AND CYS-1034, AND REVISIONS TO 1031-1032.  
RX MEDLINE-20047768; PubMed-10581026;  
RA McGuire W.T., Prasad S.D., Griffith A.J., Kunst H.P.M., Green G.E.,  
RA Shargel K.B., Runge C., Huybrechts C., Mueller R.F., Lynch E.,  
RA King M.-C., Brunner H.G., Creemers C.W.R.J., Takano M., Li S.-W.,  
RA Arita M., Wayne R., Prockop D.J., Van Camp G., Smith R.J.H.;  
RT "Mutations in COL11A2 cause non-syndromic hearing loss (DFNA13).";  
RL Nat. Genet. 23:413-419(1999).  
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY  
CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.  
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),  
CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL  
CC MODIFICATION OF ALPHA 1(XI). ALPHA 1(XI) CAN ALSO BE FOUND INSTEAD  
CC OF ALPHA 3(XI)-1(XI).  
CC -!- ALTERNATIVE PRODUCTS: 8 isoforms; 1 (shown here), 2, 3, 4, 5, 6, 7  
CC and 8; may be produced by alternative splicing. They lack exons 6,  
CC 7 or 8 or a combination of these exons.  
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -!- PTM: A DISULFIDE-BONDED PEPTIDE CALLED PROLINE/ARGININE-RICH  
CC PROTEIN OR RARP IS RELEASED FROM THE AMINO TERMINUS DURING  
CC EXTRACELLULAR PROCESSING AND IS SUBSEQUENTLY RETAINED IN THE  
CC CARTILAGE MATRIX FROM WHICH IT CAN BE ISOLATED IN SIGNIFICANT  
CC AMOUNTS.  
CC -!- DISEASE: Defects in COL11A2 are the cause of Stickler syndrome  
CC type 3 (STL3). It is an autosomal dominant disorder characterized  
CC by oro-facial, auditory and skeletal manifestations, such as  
CC midfacial hypoplasia, cleft palate, osteoarthritis, and  
CC sensorineural hearing loss. Differently from Stickler syndrome  
CC type 1 and 2, no ocular involvement is observed. This disorder is  
CC also referred to as Stickler-like syndrome or non-ocular Stickler  
CC syndrome.  
CC -!- DISEASE: Defects in COL11A2 are the cause of autosomal recessive  
CC otospondylocheirodysplasia (OSMED), a skeletal dysplasia  
CC accompanied by severe hearing loss. The phenotype overlaps that of  
CC autosomal dominant skeletal disorders (Stickler and Marshall  
CC syndromes) but can be distinguished by disproportionately short  
CC limbs and lack of ocular involvement.  
CC -!- DISEASE: Defects in COL11A2 are the cause of Weissenbacher-  
CC zweymueller syndrome (WZS), an autosomal dominant disorder allelic  
CC with STL3 and OSMED. It is also referred to as heterozygous OSMED.  
CC -!- DISEASE: Defects in COL11A2 are the cause of autosomal dominant  
CC nonsyndromic sensorineural deafness type 13 (DFNA13). Affected  
CC individuals experience progressive hearing loss beginning in the  
CC second to fourth decades, eventually making use of amplification  
CC mandatory.  
CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
CC -!- DATABASE: NAME-Hereditary hearing loss homepage;  
CC NOTE=Gene page;  
CC WWW="http://www.uia.ac.be/dnalab/hhh/hhgenes.html".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL accession -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL; U32169; AAC50214.1; -.  
CC DR EMBL; U32169; AAC50213.1; -.  
CC DR EMBL; U32169; AAC50215.1; -.  
CC DR EMBL; AL031228; CAA20240.1; -.  
CC DR EMBL; LI8987; AAA35498.1; -.  
CC DR EMBL; U04974; AAA52034.1; -.  
CC DR EMBL; U41065; AAC17464.1; -.  
CC DR EMBL; U41065; AAC17464.1; JOINED.  
CC DR EMBL; U41066; AAC17464.1; JOINED.  
CC DR EMBL; U41067; AAC17464.1; JOINED.  
CC DR PIR; A32645; A32645.  
CC DR Genew; HGNC:2187; COL11A2.  
CC DR MIM; 120290; -.  
CC DR MIM; 184840; -.  
CC DR MIM; 215150; -.  
CC DR MIM; 277610; -.  
CC DR MIM; 601868; -.  
CC DR InterPro; IPR000087; Collagen.  
CC DR InterPro; IPR000885; Fib\_collagen\_C.  
CC DR InterPro; IPR001791; Laminin\_G.  
CC DR InterPro; IPR001230; Prenyl\_site.  
CC DR InterPro; IPR003129; TSPN.  
CC DR Pfam; PF01391; Collagen; 18.  
CC DR Pfam; PF01410; COLFI; 1.  
CC DR Pfam; PF02210; TSPN; 1.  
CC DR ProDom; PD000007; Collagen; 1.  
CC DR ProDom; PD002078; Fib\_collagen\_C; 1.  
CC DR SMART; SM00038; COLFI; 1.  
CC DR SMART; SM00282; LamG; 1.  
CC DR SMART; SM00210; TSPN; 1.  
  
Query Match 62.7%; Score 47; DB 1; Length 1736;  
Best Local Similarity 80.0%; Pred. No 80;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 LGPOGRPPPP 11  
      |||||  
Db 639 LGPOGEPCPP 648  
  
RESULT 10  
ID HD\_RAT STANDARD; PRT; 3110 AA.  
AC P51111;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Huntingtin (Huntington's disease protein homolog) (HD protein).  
GN HD.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUP=Brain;  
RX MEDLINE-96133292; PubMed-8528205;  
RA Schmitt I., Baechner D., Megow D., Henklein P., Boulter J.,  
RA Hameister H., Epplen J.T., Riess O.;  
RT "Expression of the Huntington disease gene in rodents: cloning the  
RT rat homologue and evidence for downregulation in non-neuronal tissues  
RL during development.";  
RL Hum. Mol. Genet. 4:1173-1182(1995).  
RN [2]  
RP SEQUENCE OF 1773-1926 FROM N.A.  
RC TISSUP=Brain;  
RX MEDLINE=94100980; PubMed-8275091;

RA Strong T.V., Tagle D.A., Valdes J.M., Elmer L.W., Boehm K.,  
RA Swaroop M., Kaatz K.W., Collins F.S., Albin R.L.;  
RT "Widespread expression of the human and rat Huntington's disease gene  
RT in brain and nonneural tissues.";  
RL Nat. Genet. 5:259-265(1993).  
CC -!- FUNCTION: MAY PLAY A ROLE IN MICROTUBULE-MEDIATED TRANSPORT OR  
CC VESICLE FUNCTION.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- TISSUE SPECIFICITY: EXPRESSED TO A HIGH DEGREE IN ALL THE REGIONS  
CC OF THE BRAIN OF ADULTS AND IN MEIOTIC CELLS OF THE TESTIS. IN  
CC ADDITION, VERY LOW LEVELS ARE DETECTED IN VARIOUS NON-NEURONAL  
CC TISSUES (HEART, MUSCLE, LIVER, LONG AND KIDNEY).  
CC -!- DEVELOPMENTAL STAGE: IDENTIFIED AT HIGH LEVELS IN NEURONAL TISSUES  
CC OF EMBRYOS AS EARLY AS DAY 14.5. THIS EXPRESSION REMAINS CONSTANT  
CC IN ALL FURTHER DEVELOPMENT STAGES (UP TO THE ADULT). ON THE OTHER  
CC HAND THE EXPRESSION IN NON-NEURONAL TISSUES IS DOWN-REGULATED FROM  
CC STAGE 17.5 DAY OLD EMBRYOS.  
CC -!- POLYMORPHISM: THE POLY-GLN REGION DOES NOT APPEAR TO BE  
CC POLYMORPHIC, EXPLAINING THE ABSENCE OF A RODENT HD-LIKE DISORDER.  
CC -!- SIMILARITY: CONTAINS 10 HEAT REPEATS.  
CC -!- SIMILARITY: BELONGS TO THE HUNTINGTIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U18650; AAC90987.1; ALT-SEQ.  
CC EMBL: U01022; AAC52133.1; -.  
CC InterPro: IPR000091; Huntington.  
CC Pfam: PF03541; Huntington; 1.  
CC PRINTS: PR00375; HUNTINGTIN.  
CC Repeat.  
CC DOMAIN 173 297 HEAT REPEATS DOMAIN 1.  
CC DOMAIN 714 911 HEAT REPEATS DOMAIN 2.  
CC DOMAIN 1503 1544 HEAT REPEATS DOMAIN 3.  
CC DOMAIN 11 18 POLY-GLN.  
CC DOMAIN 19 38 POLY-PRO.  
CC DOMAIN 42 51 POLY-PRO.  
CC DOMAIN 1408 1411 POLY-THR.  
CC DOMAIN 2606 2611 POLY-GLU.  
CC SEQUENCE 3110 AA; 343757 MW; 33C357B8FC141550 CRC64;  
  
Query Match 62.7%; Score 47; DB 1; Length 3110;  
Best Local Similarity 87.5%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 4 PQGRPPPP 11  
Db 38 PQQPPPP 45  
  
RESULT 11  
HD\_MOUSE  
ID HD\_MOUSE STANDARD; PRT; 3119 AA.  
AC P42859.  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Huntington (Huntington's disease protein homolog) (HD protein).  
GN HD OR HDH.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).  
RC STRAIN=C57BL/6; TISSUE=Brain, and Spleen;  
RX MEDLINE=94214482; PubMed=8162057;  
RA Lin B., Nasir J., Macdonald H., Hutchinson G., Graham R.K.,

RA Rommens J.M., Hayden M.R.;  
RT "Sequence of the murine Huntington disease gene: evidence for  
RT conservation, alternate splicing and polymorphism in a triplet (CCG  
RT repeat.";  
RL Hum. Mol. Genet. 3:85-92(1994).  
CC [2]  
CC SEQUENCE FROM N.A.  
CC MEDLINE=94278649; PubMed=8003370;  
RA Barnes G.T., Duyao M.P., Ambrose C.M., McNeil S., Persichetti F.,  
RA Sinidhi J., Gusella J.F., Macdonald M.E.;  
RT "Mouse Huntington's disease gene homolog (Hdh).";  
RL Somat. Cell Mol. Genet. 20:87-97(1994).  
CC [3]  
CC SEQUENCE FROM N.A.  
CC MEDLINE=95375771; PubMed=7647777;  
RA Trottier Y., Devys D., Imbert G., Saudou F., An I., Lutz Y.,  
RA Weber C., Agid Y., Hirsch E.C., Mandel J.L.;  
RT "Cellular localization of the Huntington's disease protein and  
RT discrimination of the normal and mutated form.";  
RL Nat. Genet. 10:104-110(1995).  
CC [4]  
CC SEQUENCE OF 1-181 FROM N.A.  
CC MEDLINE=95278941; PubMed=7759106;  
RA Lin B., Nasir J., Kalchman M.A., McDonald H., Zeisler J.,  
RA Goldberg Y.P., Hayden M.R.;  
RT "Structural analysis of the 5' region of mouse and human Huntington  
RT disease genes reveals conservation of putative promoter region and  
RT dl- and trinucleotide polymorphisms.";  
RL Genomics 25:707-715(1995).  
CC -!- FUNCTION: MAY PLAY A ROLE IN MICROTUBULE-MEDIATED TRANSPORT OR  
CC VESICLE FUNCTION.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms: a long form (shown here) and a  
CC short form; seem to be produced by alternative splicing. The  
CC short form cannot be explained by a simple splicing event.  
CC -!- TISSUE SPECIFICITY: THE HIGHEST LEVEL IS SEEN THROUGHOUT THE  
CC BRAIN, BUT IT IS ALSO FOUND IN THE STOMACH, HEART, TESTIS, ADIPOSE  
CC TISSUE, MUSCLE, SPLEEN, LIVER, AND KIDNEY.  
CC -!- DEVELOPMENTAL STAGE: PREDOMINANT EXPRESSION IN NEURONAL TISSUES AT  
CC ALL DEVELOPMENTAL STAGES. IN 14.5 DAY OLD EMBRYOS, IT IS ALSO  
CC DETECTED IN NON-NEURONAL TISSUES. THIS EXPRESSION IS DOWN-  
CC REGULATED IN LATER STAGES OF DEVELOPMENT.  
CC -!- POLYMORPHISM: THE FIRST POLY-PRO REPEAT STRETCH DIFFERS IN LENGTH  
CC BY ONE UNIT (THREE) IN MUS SPRETUS STRAIN COMPARED TO OTHER  
CC SPRAINS (FOUR). THE POLY-GLN REGION DOES NOT APPEAR TO BE  
CC POLYMORPHIC, EXPLAINING THE ABSENCE OF A MURINE HD-LIKE DISORDER.  
CC -!- SIMILARITY: CONTAINS 10 HEAT REPEATS.  
CC -!- SIMILARITY: BELONGS TO THE HUNTINGTIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L23312; AAA37799.1; -.  
CC EMBL: L23313; AAA37800.1; -.  
CC EMBL: L28827; AAA89100.1; ALT\_SEQ.  
CC EMBL: U24233; AAC52218.1; -.  
CC EMBL: L34024; AAA91085.1; -.  
CC EMBL: L34008; AAA91085.1; JOINED.  
CC EMBL: L34021; AAA91085.1; JOINED.  
CC EMBL: L34022; AAA91085.1; JOINED.  
CC EMBL: L34023; AAA91085.1; JOINED.  
CC MGD: MGI:96067; Hdh.  
CC InterPro: IPR000091; Huntington.  
CC Pfam: PF03541; Huntington; 1.  
CC PRINTS: PR00375; HUNTINGTIN.  
CC Alternative splicing; Polymorphism; Repeat.  
CC DOMAIN 182 306 HEAT REPEATS DOMAIN 1.  
CC DOMAIN 723 920 HEAT REPEATS DOMAIN 2.  
CC FT

FT DOMAIN 1512 1553 HEAT REPEATS DOMAIN 3.  
FT DOMAIN 18 24 POLY-GLN.  
FT DOMAIN 25 45 POLY-PRO.  
FT DOMAIN 49 59 POLY-PRO.  
FT DOMAIN 1417 1420 POLY-THR.  
FT DOMAIN 1696 1699 POLY-GLY.  
FT DOMAIN 2615 2620 POLY-GLU.  
FT VARSPLIC MISSING (IN SHORT ISOFORM).  
FT CONFLICT 2 2 A -> G (IN REF. 1 AND 4).  
FT CONFLICT 29 29 A -> P (IN REF. 2).  
FT CONFLICT 116 116 N -> D (IN REF. 2 AND 4).  
FT CONFLICT 138 138 M -> L (IN REF. 1).  
FT CONFLICT 521 521 S -> P (IN REF. 1).  
FT CONFLICT 524 524 A -> P (IN REF. 1).  
FT CONFLICT 533 533 A -> P (IN REF. 1).  
FT CONFLICT 607 607 A -> T (IN REF. 1).  
FT CONFLICT 769 769 D -> E (IN REF. 2).  
FT CONFLICT 972 972 S -> R (IN REF. 1).  
FT CONFLICT 1106 1106 W -> C (IN REF. 1).  
FT CONFLICT 1240 1240 T -> N (IN REF. 1).  
FT CONFLICT 1384 1384 N -> T (IN REF. 1).  
FT CONFLICT 1827 1827 H -> Y (IN REF. 1).  
FT CONFLICT 1979 1980 PF -> SS (IN REF. 1).  
FT CONFLICT 2062 2062 D -> G (IN REF. 1).  
FT CONFLICT 2570 2570 S -> N (IN REF. 1).  
FT CONFLICT 2866 2866 E -> V (IN REF. 1).  
FT CONFLICT 2877 2877 V -> G (IN REF. 1).  
FT CONFLICT 2882 2882 D -> G (IN REF. 1).  
FT CONFLICT 2887 2887 Q -> H (IN REF. 1).  
FT CONFLICT 2915 2915 A -> T (IN REF. 1).  
FT CONFLICT 3025 3025 P -> S (IN REF. 3).  
FT CONFLICT 3062 3063 QV -> LM (IN REF. 1).  
FT CONFLICT 3095 3096 WV -> EE (IN REF. 1).  
SQ SEQUENCE 3119 AA; 344685 MW; ECA42B5916F50F4F CRC64;  
  
Query Match 62.78; Score 47; DB 1; Length 3119;  
Best Local Similarity 87.5%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 4 PQGRPPPP 11  
Db 45 PQGQPPPP 52  
|||||  
  
RESULT 12  
BTBL\_HUMAN  
ID BTBL\_HUMAN STANDARD; PRT; 482 AA.  
AC Q9H0C5; Q9BX71; Q9WN4;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE BTB/POZ domain containing protein 1.  
GN BTBD1 OR C15ORF1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.  
RX MEDLINE=21100466; PubMed=11179693;  
RA Carim-Todd L., Sumoy L., Andreu N., Estivill X., Escarceller M.;  
RT "Identification and characterization of BTBD1, a novel BTB domain  
containing gene on human chromosome 15q24.";  
RL Gene 262:275-281(2001).  
RN [2]  
RN SEQUENCE FROM N.A.  
RA Yang L., Xu L., D'Arpa P.;  
RT "Characterization of two similar BTB-containing Kelch-like proteins,  
BTBD1 and BTBD2, identified as topoisomerase I-interacting  
proteins.";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RN SEQUENCE FROM N.A.

RC TISSUE=Testis;  
RX MEDLINE=21154917; PubMed=11230166;  
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,  
RA Ansorge W., Boecker M., Bloecker H., Bauersachs S., Blum H.,  
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,  
RA Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,  
RA Wambutt R., Korn B., Klein M., Poustka A.;  
RT "Towards a catalog of human genes and proteins: sequencing and  
analysis of 500 novel complete protein coding human cDNAs.";  
RL Genome Res. 11:422-435(2001).  
RN [4]  
RN SEQUENCE FROM N.A.  
RP Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,  
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBUNIT: Interacts with topoisomerase 1.  
CC -!- TISSUE SPECIFICITY: Ubiquitous; higher levels in heart and  
skeletal muscle.  
CC -!- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AL136853; CAB66787.1; -  
DR EMBL; AF355402; AAK25825.1; -  
DR EMBL; AF257241; AAK17068.1; -  
DR EMBL; AK000731; BAA91345.1; -  
DR Genew; HGNC:1120; BTBD1.  
DR InterPro; IPR000210; BTB\_POZ.  
DR Pfam; PF00651; BTB; 1.  
DR SMART; SM00225; BTB; 1.  
DR PROSITE; PS50097; BTB; 1.  
FT DOMAIN 69 145 BTB.  
FT DOMAIN 25 31 POLY-PRO.  
FT CONFLICT 406 406 T -> A (IN REF. 4).  
FT CONFLICT 422 422 C -> G (IN REF. 2).  
FT CONFLICT 429 429 L -> P (IN REF. 4).  
SQ SEQUENCE 482 AA; 52771 MW; 525A49E01728AFF0 CRC64;  
  
Query Match 61.3%; Score 46; DB 1; Length 482;  
Best Local Similarity 72.7%; Pred. No. 34;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 ELGPGQGRPPPP 11  
Db 19 EPGAGPPPPPP 29  
|||||  
  
RESULT 13  
VINC\_CAEEL  
ID VINC\_CAEEL STANDARD; PRT; 1010 AA.  
AC P19826;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Vinculin (P107B).  
GN DEB-1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=Bristol N2.  
RX MEDLINE=89255506; PubMed=2498337;  
RA Barstead R.J., Waterston R.H.;

RT "The basal component of the nematode dense-body is vinculin.";

RL J. Biol. Chem. 264:10177-10185(1989).

CC -!- FUNCTION: INVOLVED IN CELL ADHESION. MAY BE INVOLVED IN THE

CC ATTACHMENT OF THE ACTIN-BASED MICROFILAMENTS TO THE PLASMA

CC MEMBRANE.

CC -!- SUBUNIT: EXHIBITS SELF-ASSOCIATION PROPERTIES.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF ADHESION PLAQUES.

CC -!- SIMILARITY: STRONG. TO OTHER VINCULINS AND TO ALPHA-CATENINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; J04804; AAA28002.1; -

DR PIR; A33509; A33509.

DR InterPro; IPR001043; Vinculin/catenin.

DR InterPro; IPR000633; Vinculin\_2.

DR Pfam; PF01044; Vinculin; 1.

DR PRINTS; PR00806; VINCULIN.

DR ProDom; PD025583; VINCULIN\_2; 2.

DR PROSITE; PS00663; VINCULIN\_1; 1.

DR PROSITE; PS00664; VINCULIN\_2; 2.

KW Cell adhesion; Actin-binding; Cytoskeleton; Structural protein;

KW Phosphorylation; Repeat.

FT DOMAIN 259 481 2 X REPEATS.

FT REPEAT 259 366 1.

FT REPEAT 375 481 2.

FT DOMAIN 786 807 PRO-RICH.

FT SIMILAR 471 490 TO SOYBEAN ACTIN (AA 178-197).

SQ SEQUENCE 1010 AA; 111922 MW; DAF789EA4C786721 CRC64;

Query Match 61.3%; Score 46; DB 1; Length 1010;

Best Local Similarity 72.7%; Pred. No. 66;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ELGPOGRPPPP 11

DB 796 ELSPPPRPPPP 806

RESULT 14

HXA2\_CHICK

ID HXA2\_CHICK STANDARD; PRT; 375 AA.

AC Q08727;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Homeobox protein Hox-A2.

DE HOXA2 OR HOXA-2.

GN Gallus gallus (Chicken).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95324377; PubMed=7600967;

RA Prince V., Lumsden A.;

RT "Hoxa-2 expression in normal and transposed rhombomeres: independent

RT regulation in the neural tube and neural crest.";

RL Development 120:911-923(1994).

CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF

CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH

CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.

CC -!- PROBOSCIPEDIA SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X74323; CA552370.1; -

DR HSSP; P14533; 1B72.

DR TRANSFAC; T01700; -

DR InterPro; IPR001827; Antennapedia.

DR InterPro; IPR001356; Homeobox.

DR Pfam; PF00046; homeobox; 1.

DR PRINTS; PR00025; ANTENNAPEDIA.

DR PRINTS; PR00024; HOMEBOX.

DR ProDom; PD000010; Homeobox; 1.

DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX\_1; 1.

DR PROSITE; PS00071; HOMEBOX\_2; 1.

DR PROSITE; PS00032; ANTENNAPEDIA; 1.

KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;

KW Transcription regulation.

FT DOMAIN 94 99 ANTP-TYPE HEXAPEPTIDE.

FT DNA\_BIND 140 199 HOMEBOX.

SQ SEQUENCE 375 AA; 40998 MW; C7C4CA10455C3B11 CRC64;

Query Match 60.0%; Score 45; DB 1; Length 375;

Best Local Similarity 77.8%; Pred. No. 36;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GPOGRPPPP 11

DB 85 GPACAPPPP 93

RESULT 15

HIC2\_HUMAN

ID HIC2\_HUMAN STANDARD; PRT; 615 AA.

AC Q96JB3; Q9UPX9; Q9NSM9; Q96KR3;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypermethylated in cancer 2 protein (Hic-2) (Hic-3) (HIC1-related gene

DE on chromosome 22).

GN HIC2 OR HRG22 OR KIAA1020.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE OF 1-85 FROM N.A. (ISOFORM 1), AND ALTERNATIVE SPLICING.

RX MEDLINE=21438979; PubMed=11554746;

RA Deltour S., Pinte S., Guerardel C., Leprince D.;

RT "Characterization of HRG22, a human homologue of the putative tumor

RT suppressor gene HIC1.";

RL Biochem. Biophys. Res. Commun. 287:427-434(2001).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM 2).

RC TISSUE=Brain;

RA Alliel P.M., Goudou D., Bitoun M., Seddiqi N., Rieger F., Perin J.-P.;

RT "Complete deduced structure of HIC-3, a novel human btb/poz and zf

RT factor of the HIC family.";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=20057165; PubMed=10591208;

RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,

RA Clamp M., Smith L.J., Alnsough R., Almeida J.P., Babbage A.K.,

RA Bagquley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,

RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,

RA Burdill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,

RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,

RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,

RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
RA Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,  
RA Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,  
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,  
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,  
RA McCall J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,  
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,  
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,  
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,  
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,  
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,  
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,  
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,  
RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,  
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,  
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,  
RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,  
RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,  
RA Serousi E., Franssion I., Tapia I., Bruder C.E., O'Brien K.P.,  
RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,  
RA Tilahun Y., Wright H.;  
RL "The DNA sequence of human chromosome 22.";  
RL Nature 402:489-495(1999).  
RN [4]  
RP SEQUENCE OF 61-615 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=93397452; PubMed=10470851;  
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,  
RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XIV.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 6:197-205(1999).  
RN [5]  
RP SEQUENCE OF 301-615 FROM N.A.  
RC TISSUE=Testis;  
RA Koehler K., Beyer A., Mewes H.-W., Weil B., Wiemann S.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Transcriptional repressor.  
CC -!- SUBUNIT: Interacts with CTBP.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
CC produced by alternative splicing.  
CC -!- TISSUE SPECIFICITY: Highest levels in cerebellum.  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS. HIC TRANSCRIPTION FACTORS SUBFAMILY.  
CC -!- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AJ313204; CAC70715.1; -;  
DR EMBL; AF349035; AAK72951.1; -;  
DR EMBL; AP000557; -; NOT-ANNOTATED\_CDS.  
DR EMBL; AB028943; BAA82972.1; -;

DR EMBL; AL162003; CAB82344.1; -;  
DR Genew; HGNC:18595; HIC2.  
DR InterPro; IPR000210; BTB\_POZ.  
DR InterPro; IPR000822; Znf\_C2H2.  
DR Pfam; PF00651; BTB; 1.  
DR Pfam; PF00096; zf-C2H2; 5.  
DR SMART; SM00355; Znf\_C2H2; 5.  
DR PROSITE; PS00097; BTB; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 5.  
KW Multigene family; Nuclear protein; Transcription regulation;  
KW DNA-binding; Zinc-finger; Metal-binding; Repeat; Repressor;  
KW Alternative splicing.  
FT DOMAIN 46 109 BTB.  
FT ZN\_FING 442 615 ZINC FINGERS.  
FT ZN\_FING 442 459 C2H2-TYPE.  
FT ZN\_FING 505 532 C2H2-TYPE.  
FT ZN\_FING 533 560 C2H2-TYPE.  
FT ZN\_FING 561 588 C2H2-TYPE.  
FT SITE 246 250 BINDING TO CTBP.  
FT VARSPLIC 1 18 MISSING (IN ISOFORM 2).  
FT CONFLICT 49 51 IIM -> TIR (IN REF. 1).  
FT CONFLICT 176 176 Q -> R (IN REF. 4).  
SQ SEQUENCE 615 AA; 66155 MW; B0368C631B198C95 CRC64;  
  
Query Match 60.0%; Score 45; DB 1; Length 615;  
Best Local Similarity 70.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 3 GPGQGRPPPC 12  
||| ||| ||  
Db 386 GPGYEPYPYC 395

Search completed: February 5, 2003, 09:48:19  
Job time : 10.5714 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:45:51 : Search time 28 Seconds  
(without alignments)  
88.306 Million cell updates/sec

Title: US-09-700-329-2

Perfect score: 75

Sequence: 1 ELGPGGRPPPPC 12

Scoring table:

BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	68.0	261	10	09CAE4
2	51	68.0	297	11	09JLV7
3	50	66.7	411	16	09RCM3
4	49	65.3	858	13	08UXX0
5	49	65.3	2241	5	015850
6	48	64.0	211	5	045098
7	48	64.0	327	5	001799
8	48	64.0	373	4	09NP08
9	48	64.0	1616	4	015054
10	47	62.7	67	11	09Z1I8
11	47	62.7	329	12	009798
12	47	62.7	333	16	09L009
13	47	62.7	635	12	096087
14	47	62.7	1827	13	08UUM5
15	46	61.3	112	17	09YER0
16	46	61.3	139	11	09CTN9

17	46	61.3	230	5	09W3V7	09W3V7 drosophila
18	46	61.3	373	5	09B131	09B131 caenorhabd
19	46	61.3	502	10	08S5L3	08S5L3 oryza sativ
20	46	61.3	579	10	09LGG8	09LGG8 oryza sativ
21	46	61.3	993	5	09YV31	09YV31 drosophila
22	46	61.3	1014	5	09B132	09B132 caenorhabd
23	46	61.3	1500	10	094DDA0	094DDA0 oryza sativ
24	46	61.3	1784	10	08S5L6	08S5L6 oryza sativ
25	45	60.0	253	5	09VME9	09VME9 drosophila
26	45	60.0	363	10	09ZU66	09ZU66 arbidopsis
27	45	60.0	363	10	094C63	094C63 arbidopsis
28	45	60.0	388	4	096H72	096H72 homo sapien
29	45	60.0	432	5	021414	021414 caenorhabd
30	45	60.0	533	10	08VYS5	08VYS5 arbidopsis
31	45	60.0	533	10	09STK1	09STK1 arbidopsis
32	45	60.0	705	4	08TER5	08TER5 homo sapien
33	45	60.0	756	16	P72877	P72877 synechocyst
34	45	60.0	1184	5	09V5D2	09V5D2 drosophila
35	45	60.0	1820	13	091907	091907 pagrus majo
36	44	58.7	85	11	070555	070555 mus musculu
37	44	58.7	142	11	08VCS6	08VCS6 mus musculu
38	44	58.7	142	11	08R5H4	08R5H4 mus musculu
39	44	58.7	173	4	015214	015214 homo sapien
40	44	58.7	176	10	09C4Z8	09C4Z8 arbidopsis
41	44	58.7	238	4	000600	000600 homo sapien
42	44	58.7	243	2	09F104	09F104 thermus the
43	44	58.7	297	4	016038	016038 homo sapien
44	44	58.7	309	4	004118	004118 homo sapien
45	44	58.7	329	5	017166	017166 brugia mala

## ALIGNMENTS

RESULT 1  
ID 09CAE4 PRELIMINARY; PRT; 261 AA.  
AC 09CAE4;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hypothetical 29.4 kDa protein.  
GN F13M14.33.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=21016720; PubMed=11130713;  
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseid M.,  
RA Farmanan B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,  
RA Delzeny M., Boutry M., Grivell L.A., Mache R., Puidomenech P.,  
RA De Simone V., Choise N., Artiguenave F., Robert C., Brotier P.,  
RA Wincker P., Catolico L., Weissenbach J., Saurin W., Quetier F.,  
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,  
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,  
RA Vezzi A., D'Angelo M., Pallavicini A., Toppi S., Simonati B.,  
RA Conrad A., Hornischer K., Kauer G., Loehert T.-H., Nordsiek G.,  
RA Reichelt J., Scharfe M., Schoen O., Barques M., Terol J., Climent J.,  
RA Navarro P., Collado C., Perez-Perez A., Ottenwalder B., Duchemin D.,  
RA Cooke R., Laurie M., Berger-Liandro C., Purnelle B., Masuy D.,  
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,  
RA Monfort A., Argitlou A., Flores M., Liguori R., Vitale D.,  
RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,  
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
RA Rooney T., Rizzo M., Walts A., Uitterback T., Fujii C.Y., Shea T.P.,  
RA Creasy T.H., Haas B., Marti R., Wu D., Peterson J., Van Aken S.,  
RA Pal G., Miltner J., Sellars P., Gill J.E., Feldblum T.V.,  
RA Preuss D., Lin X., Niernan W.C., Salzberg S.L., White O., Venter J.C.,

Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
 Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,  
 Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 Nakayama S., Nakazaki N., Shinjo S., Takeuchi C., Wada T.,  
 Watanabe A., Yamada M., Yasuda M., Tabata S.;  
 "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
 thaliana";  
 RL Nature 408:820-822(2000).  
 DR EMBL: AC011560; AAG51392.1; -.  
 DR HSSP: P19339; ISXL.  
 DR InterPro: IPR000504; RNA\_rec\_mot.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR Pfam: PF00076; rtm: 1.  
 DR Pfam: PF00098; zf-CCHC; 1.  
 DR SMART: SM00360; RRM\_1.  
 DR SMART: SM00343; Znf\_C2HC; 1.  
 DR PROSITE: PS50102; RRM; 1.  
 DR PROSITE: PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 261 AA; 29367 MW; 4A37767C92819D22 CRC64;

Query Match 68.0%; Score 51; DB 10; Length 261;  
 Best Local Similarity 72.7%; Pred. No. 3.8;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 ELPGGRRPP 11  
 DB 171 QLGERRRPP 181

## RESULT 2

O9JLV7 PRELIMINARY; PRT; 297 AA.

AC 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Lung kruppel-like factor (Fragment).  
 GN KLF2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=128;  
 RX MEDLINE=99365315; PubMed=10433980;  
 RA Schrick J.J., Hughes M.J., Anderson K.P., Croyle M.L., Lingrel J.B.;  
 RT "Characterization of the lung Kruppel-like transcription factor gene  
 RT and upstream regulatory elements";  
 RL Gene 236:185-195(1999)  
 DR EMBL: AF129002; AAR33702.1; -.  
 DR MGD: MGI:1342772; Klf2.  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR InterPro: IPR000822; Znf\_C2H2.  
 DR Pfam: PF00096; zf-C2H2; 1.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 DR SMART: SM00355; Znf\_C2H2; 1.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1;  
 DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 1.  
 KW DNA-binding; Metal-binding; Zinc-finger.  
 FT NON\_TER 297  
 SQ SEQUENCE 297 AA; 30812 MW; 4DF288DE1FE65E1D CRC64;

Query Match 68.0%; Score 51; DB 11; Length 297;  
 Best Local Similarity 88.9%; Pred. No. 4.3;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GPGGRRPP 11  
 DB 160 GPAGRRPP 168

RESULT 3  
 O9FCM3 PRELIMINARY; PRT; 411 AA.  
 AC O9FCM3;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative secreted oxidoreductase (Fragment).  
 GN SC07791A OR SC10B8A.01C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_Taxid=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Brown S.P., Harris D.;  
 RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;  
 RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,  
 RA Kinash H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2)";  
 RL Nature 417:141-147(2002).  
 DR EMBL: AL391454; CAC04103.1; -.  
 DR InterPro: IPR000733; Flav\_monooxygenase.  
 DR InterPro: IPR003042; Rng\_monooxygenase.  
 DR Pfam: PF01360; Monooxygenase; 1.  
 DR PRINTS: PR00420; RNMNOXGNASE.  
 DR NON\_TER 411  
 FT NON\_TER 411  
 SQ SEQUENCE 411 AA; 44447 MW; 7B994CC54E8A3E86 CRC64;

Query Match 66.7%; Score 50; DB 16; Length 411;  
 Best Local Similarity 100.0%; Pred. No. 8.1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 GPGGRRPP 10  
 DB 380 GPAGRRPP 387

RESULT 4  
 O8UVX0 PRELIMINARY; PRT; 858 AA.  
 AC O8UVX0;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE P1wi protein.  
 OS Brachydanio rerio (zebrafish) (zebra danio).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN
  [1]
  RP SEQUENCE FROM N.A.
  RA Weeraratne S.D., Gong Z., Tan C.-H.;
  RT "Cloning and characterization of zebrafish homolog of p1w1, essential
  RT for germ-line stem cell self-renewal.";
  RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF336369; AAL51170.1; -.
DR InterPro: IPR003100; PAZ.
DR Pfam: PF02170; PAZ; 1.
DR Pfam: PF02171; P1w1; 1.
SQ SEQUENCE 858 AA; 97451 MW; 6A12F2E511465777 CRC64;

Query Match
Best Local Similarity 65.3%; Score 49; DB 13; Length 858;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LGPGRPP 10
Db 366 LGPGSRPP 374

RESULT 5
015850 PRELIMINARY; PRT; 2241 AA.
AC 015850;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE L3162.1.
GN L3162.1.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN
  [1]
  RP SEQUENCE FROM N.A.
  RC STRAIN-FRIEDLIN;
  RX MEDLINE=99178987; PubMed=10077609;
  RA Myler P.J., Audleman L., devos T., Hixson G., Kiser P., Lemley C.,
  RA Magness C., Rickel E., Sisk E., Sunkin S., Swartzell S., Westlake T.,
  RA Bastien P., Fu G., Ivens A., Stuart K.;
  RT "Leishmania major Friedlin chromosome 1 has an unusual distribution of
  RT protein-coding genes.";
  RL Proc. Natl. Acad. Sci. U.S.A. 96:2902-2906(1999).
DR EMBL: AE001274; AAC24680.1; -.
DR InterPro: IPR000008; C2.
DR InterPro: IPR000937; Viral coat.
DR PRINTS: PR00233; ICOSAHEDRAL.
DR SMART: SM00233; C2; 1.
DR PROSITE: PS50004; C2_DOMAIN_2; 1.
SQ SEQUENCE 2241 AA; 241669 MW; 7751B207A5755421 CRC64;

Query Match
Best Local Similarity 65.3%; Score 49; DB 5; Length 2241;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 PGRPPPP 12
Db 1855 PGRGPPPP 1863

RESULT 6
045098 PRELIMINARY; PRT; 211 AA.
AC 045098;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 24.1 kDa protein.

```

```

GN F57H12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderrinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN
  [1]
  RP SEQUENCE FROM N.A.
  RC STRAIN-BRISTOL N2;
  RX MEDLINE=99069613; PubMed=9851916;
  RA None;
  RT "Genome sequence of the nematode C. elegans: a platform for
  RT investigating biology. The C. elegans Sequencing Consortium.";
  RL Science 282:2012-2018(1998).
RN
  [2]
  RP SEQUENCE FROM N.A.
  RC STRAIN-BRISTOL N2;
  RA Bentley D., Le T.T.;
  RT "The sequence of C. elegans cosmid F57H12.";
  RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN
  [3]
  RP SEQUENCE FROM N.A.
  RC STRAIN-BRISTOL N2;
  RA Waterston R.;
  RT "Direct Submission.";
  RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF045644; AAC02599.1; -.
DR InterPro: IPR003677; Onchocerca_Ag.
DR Pfam: PF02520; DUF148; 1.
DR Hypothetical protein.
SQ SEQUENCE 211 AA; 24129 MW; DD8B799E19DD26F3 CRC64;

Query Match
Best Local Similarity 64.0%; Score 48; DB 5; Length 211;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GRPPPP 12
Db 30 GRPPPP 36

RESULT 7
001799 PRELIMINARY; PRT; 327 AA.
AC 001799;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 33.0 kDa protein.
GN F53G12.7
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderrinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN
  [1]
  RP SEQUENCE FROM N.A.
  RC STRAIN-BRISTOL N2;
  RX MEDLINE=99069613; PubMed=9851916;
  RA None;
  RT "Genome sequence of the nematode C. elegans: a platform for
  RT investigating biology. The C. elegans Sequencing Consortium.";
  RL Science 282:2012-2018(1998).
RN
  [2]
  RP SEQUENCE FROM N.A.
  RC STRAIN-BRISTOL N2;
  RA Wu X., Graves T.;
  RT "The sequence of C. elegans cosmid F53G12.";
  RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN
  [3]
  RP SEQUENCE FROM N.A.
  RC STRAIN-BRISTOL N2;
  RA Waterston R.;
  RT "Direct Submission.";
  RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

```

EMBL: AF003139; AAB54156.1; -  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR002486; COL\_cuticle\_N.  
 DR Pfam: PF01391; Collagen; 2.  
 DR Pfam: PF01484; COL\_cuticle\_N; 1.  
 DR Prodom: PD000007; Collagen; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 327 AA; 33025 MW; 1DFACB590408590C CRC64;

Query Match 64.0%; Score 48; DB 5; Length 327;  
 Best Local Similarity 72.7%; Pred. No. 13;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EIGPGRPPPP 11  
 DB 260 EMGPGRPPPP 270

## RESULT 8

Q9NP08 PRELIMINARY; PRT; 373 AA.  
 AC Q9NP08; 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DE H6 homeodomain protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=CRANIOFACIAL;  
 RA MEDLINE=93087572; Pubmed=1360670;  
 RA Stadler H.S., Padanilam B.J., Buetow K., Murray J.C., Solursh M.;  
 RT "Identification and genetic mapping of a homeobox gene to the 4p16.1  
 RT region of human chromosome 4.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:11579-11583(1992).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL: M99587; AAF70205.1; -;  
 DR HSSP: P14653; 1872.  
 DR TRANSFAC: T04443; -;  
 DR InterPro: IPR001356; Homeobox.  
 DR InterPro: IPR000047; HTH\_repressr.  
 DR Pfam: PF000046; homeobox; 1.  
 DR PRINTS: PR00024; HOMEBOX.  
 DR PRINTS: PR00031; HTHREPRESSR.  
 DR PRODOM: PD000010; Homeobox; 1.  
 DR SMART: SM00389; HOX; 1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 KW DNA-binding; Homeobox; Nuclear protein.  
 SQ SEQUENCE 373 AA; 39225 MW; 8BF9EB1722660A76 CRC64;

Query Match 64.0%; Score 48; DB 4; Length 373;  
 Best Local Similarity 77.8%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 PGRPPPP 12  
 DB 325 PPSRPPPP 333

## RESULT 9

Q15054 PRELIMINARY; PRT; 1616 AA.  
 AC Q15054; 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE KIA0346 protein (Fragment).  
 GN KIAA0346.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RA MEDLINE=97349984; Pubmed=9205841;  
 RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,  
 RA Tanaka A., Kori H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 4:141-150(1997).  
 DR EMBL: AB002344; BAA21572.1; -;  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR InterPro: IPR003347; TF\_jmjc.  
 DR Pfam: PF02373; Jmjc; 1.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 FT NON\_TER 1  
 SQ SEQUENCE 1616 AA; 174114 MW; B571D366792E8B9C CRC64;

Query Match 64.0%; Score 48; DB 4; Length 1616;  
 Best Local Similarity 70.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GPGRPPPP 12  
 DB 398 GPSSPPPP 407

## RESULT 10

Q92118 PRELIMINARY; PRT; 67 AA.  
 AC Q92118; 01-MAY-1999 (TRENBLrel. 10, Created)  
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
 DE Huntingtin (Fragment).  
 GN RH.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SPLEN;  
 RA MEDLINE=99025916; Pubmed=9806905;  
 RA Holzmann C., Maeueller W., Petersohn D., Schmidt T., Thiel G.,  
 RA Epplen J.T., Rless O.;  
 RT "Isolation and characterization of the rat Huntingtin promoter.";  
 RL Biochem. J. 336:227-234(1998).  
 DR EMBL: AJ224997; CA11281.1; -;  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 FT NON\_TER 67  
 SQ SEQUENCE 67 AA; 7296 MW; F0AA9C008D7FDD7B CRC64;

Query Match 62.7%; Score 47; DB 11; Length 67;  
 Best Local Similarity 87.5%; Pred. No. 4.4;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 PGRPPPP 11  
 DB 46 PGOPPPPP 53

## RESULT 11

Q09798 PRELIMINARY; PRT; 329 AA.  
 AC Q09798; 01-JUL-1997 (TRENBLrel. 04, Created)  
 DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE UL26.5 protein.  
 GN UL26.5.  
 OS human herpesvirus 1.  
 OC Viruses: dsDNA viruses, no RNA stage: Herpesviridae;  
 OC Alphaherpesvirinae: Simplexvirus.  
 OX NCBI\_TaxID=10298;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17;  
 RA McGeoch D.J.;  
 RL Submitted (JAN-1989) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17;  
 RA MEDLINE=88274327; PubMed=2839594;  
 RX McGeoch D.J., Dairymple M.A., Davison A.J., Dolan A., Frame M.C.,  
 RA McNab D., Perry L.J., Scott J.E., Taylor P.;  
 RT "The complete DNA sequence of the long unique region in the genome of  
 RT herpes simplex virus type 1.";  
 RL J. Gen. Virol. 69:1531-1574(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17;  
 RA MEDLINE=86148504; PubMed=3005980;  
 RX McGeoch D.J., Dolan A., Donald S., Brauer D.H.;  
 RT "Complete DNA sequence of the short repeat region in the genome of  
 RT herpes simplex virus type 1.";  
 RL Nucleic Acids Res. 14:1727-1745(1986).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17;  
 RA MEDLINE=92341080; PubMed=1321882;  
 RX Dolan A., McKie E., Maclean A.R., McGeoch D.J.;  
 RT "Status of the ICP34.5 gene in herpes simplex virus type 1 strain  
 RT 17.";  
 RL J. Gen. Virol. 73:971-973(1992).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17;  
 RA Dolan A.;  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: X14112; CAA32319.1; -;  
 DR InterPro: IPR001847; Assemblin.  
 DR Pfam: PF00716; Peptidase\_S21; 1.  
 SQ SEQUENCE 329 AA; 33760 MW; DF690388D0630F9D CRC64;  
  
 Query Match 62.7%; Score 47; DB 12; Length 329;  
 Best Local Similarity 77.8%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
 OY 4 PGGPPPPC 12  
 DB 259 PGGPPPPC 267  
  
 RESULT 12  
 Q9L009 PRELIMINARY; PRT; 333 AA.  
 AC Q9L009;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Putative dehydrogenase.  
 GN SC02304 OR SCC30.12C.  
 OS Streptomyces coelicolor.  
 OC Bacteria: Firmicutes: Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger K.J., Harris D.;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Godle A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2)."  
 RL Nature 417:141-147(2002).  
 DR EMBL: ALJ52972; CAB88182.1; -;  
 DR InterPro: IPR002539; MacC\_dehydratas.  
 DR Pfam: PF01575; MacC\_dehydratas; 1.  
 SQ SEQUENCE 333 AA; 35241 MW; FE07C1CC04BF612D CRC64;

Query Match 62.7%; Score 47; DB 16; Length 333;  
 Best Local Similarity 72.7%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ELGGPPPP 11  
 DB 188 ERGSGRPP 198  
  
 RESULT 13  
 Q69087 PRELIMINARY; PRT; 635 AA.  
 AC Q69087;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE UL26.  
 OS human herpesvirus 1.  
 OC Viruses: dsDNA viruses, no RNA stage: Herpesviridae;  
 OC Alphaherpesvirinae: Simplexvirus.  
 OX NCBI\_TaxID=10298;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=94308117; PubMed=8034621;  
 RX Darke P.L., Chen E., Hall D.L., Sardana M.K., Veloski C.A.,  
 RA LaFemina R.L., Shafer J.A., Kuo L.C.;  
 RT "Purification of active Herpes simplex virus-1 protease expressed in  
 RT Escherichia coli.";  
 RL J. Biol. Chem. 269:18708-18711(1994).  
 DR EMBL: L32018; AAA45828.1; -;  
 DR HSSP: P16753; IWPO.  
 DR MEROPS: S21.001; -;  
 DR InterPro: IPR001847; Assemblin.  
 DR Pfam: PF00716; Peptidase\_S21; 1.  
 DR PRINTS: PR00236; HSCVAPSIDP40.  
 FT CHAIN 1 247 HSV-1 PROTEASE.  
 FT CHAIN 253 610 ICP35 ASSEMBLY PROTEIN.  
 SQ SEQUENCE 635 AA; 66454 MW; 3C7768BE1496FDBF CRC64;

Query Match 62.7%; Score 47; DB 12; Length 635;  
 Best Local Similarity 77.8%; Pred. No. 34;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 PGRPPPC 12  
 11111111  
 DB 565 PGRPPPC 573

## RESULT 14

ID Q8UM5 PRELIMINARY; PRT; 1827 AA.  
 AC Q8UM5;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE COL11A2 protein.  
 GN COL11A2.  
 OS Oryzias latipes (Medaka fish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Belontiiformes; Adrianiichthyidae; Oryziinae; Oryzias.  
 OX NCBI\_TaxID=8090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HD-RR.  
 RA Matsuo M.Y., Asakawa S., Shimizu N., Kimura H., Nonaka M.;  
 RT "Nucleotide sequence of the MHC Class I Region of a Teleost, the  
 Medaka.";  
 RL Immunogenetics 0:0-0(2002).  
 DR EMBL: AB073376; BAB83839.1; -;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib-collagen\_C.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR InterPro: IPR003129; TSPN.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR Pfam: PF02210; TSPN; 1.  
 DR ProDom: PD000007; Collagen; 3.  
 DR ProDom: PD002078; Fib-collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 DR SMART: SM00282; LamG; 1.  
 DR SMART: SM00210; TSPN; 1.  
 DR PROSITE: PS50025; LAM\_G\_DOMAIN; 1.  
 SQ SEQUENCE 1827 AA; 181589 MW; AF5C71E9B38906EA CRC64;

Query Match 62.7%; Score 47; DB 13; Length 1827;  
 Best Local Similarity 80.0%; Pred. No. 88;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 LGPGRPPPP 11  
 11111111  
 DB 719 LGPGRPPPP 728

## RESULT 15

ID Q9YEB0 PRELIMINARY; PRT; 112 AA.  
 AC Q9YEB0;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical protein APE0664.  
 GN APE0664.  
 OS Aeropyrum pernix.  
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;  
 OC Desulfurococcaceae; Aeropyrum.  
 OX NCBI\_TaxID=56636;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KL;

RX MEDLINE=99310339; PubMed=10382966;  
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,  
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,  
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic hyper-thermophilic  
 crenarchaeon, Aeropyrum pernix K1.";  
 RL DNA Res. 6:83-101(1999).  
 DR EMBL: AP000060; BAA79636.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 112 AA; 11931 MW; 9007CDD89321151E CRC64;

Query Match 61.3%; Score 46; DB 17; Length 112;  
 Best Local Similarity 77.8%; Pred. No. 9.8;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 PGRPPPC 12  
 11111111  
 DB 71 PGRPPPC 79

Search completed: February 5, 2003, 09:49:16  
 Job time : 30 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:46:51 ; Search time 13.1429 Seconds  
(without alignments)  
26.864 Million cell updates/sec

Title: US-09-700-329-2

Perfect score: 75

Sequence: 1 ELGPGGRPPPPC 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents, AA:\*
- 1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*
  - 2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*
  - 3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*
  - 4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*
  - 5: /cgn2\_6/ptodata/1/1aa/PCUTS.COMB.pep:\*
  - 6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	74.7	11	1	US-08-446-692-73 Sequence 73, Appli
2	56	74.7	11	2	US-08-488-351A-73 Sequence 73, Appli
3	48	64.0	16	1	US-08-188-223-2 Sequence 2, Appli
4	48	64.0	16	4	US-08-968-466-2 Sequence 2, Appli
5	48	64.0	16	4	US-08-478-546B-2 Sequence 2, Appli
6	47	62.7	635	1	US-07-832-855-2 Sequence 2, Appli
7	47	62.7	635	4	US-08-176-320-2 Sequence 2, Appli
8	47	62.7	3118	2	US-08-457-273B-8 Sequence 8, Appli
9	47	62.7	3119	1	US-08-246-982A-16 Sequence 16, Appli
10	47	62.7	3119	1	US-08-453-265-16 Sequence 16, Appli
11	46	61.3	878	1	US-08-237-919-2 Sequence 2, Appli
12	46	61.3	878	4	US-08-732-428-2 Sequence 2, Appli
13	46	61.3	878	4	US-09-798-267-3 Sequence 2, Appli
14	46	61.3	878	4	US-09-798-267-3 Sequence 2, Appli
15	46	61.3	878	4	US-09-798-267-3 Sequence 2, Appli
16	45	60.0	43	4	US-08-908-371B-12 Sequence 12, Appli
17	45	60.0	659	4	US-09-562-737-14 Sequence 14, Appli
18	44	58.7	82	2	US-08-925-237-4 Sequence 14, Appli
19	44	58.7	90	4	US-09-314-268-136 Sequence 136, App
20	44	58.7	124	4	US-08-925-237-2 Sequence 2, Appli
21	44	58.7	322	4	US-09-383-586-33 Sequence 33, Appli
22	43.5	58.0	1122	2	US-08-619-198-3 Sequence 3, Appli
23	43.5	58.0	1389	2	US-08-619-198-5 Sequence 5, Appli
24	43	57.3	15	1	US-08-151-219-4 Sequence 4, Appli
25	43	57.3	15	5	PCR-US94-13205-4 Sequence 4, Appli
26	43	57.3	17	1	US-08-188-223-6 Sequence 6, Appli
27	43	57.3	17	4	US-08-968-466-6 Sequence 6, Appli

28	43	57.3	17	4	US-08-478-546B-6 Sequence 6, Appli
29	43	57.3	446	3	US-08-956-254-2 Sequence 2, Appli
30	43	57.3	446	3	US-09-008-388-1 Sequence 1, Appli
31	43	57.3	448	2	US-09-015-815-1 Sequence 1, Appli
32	42	56.0	6	1	US-08-151-219-5 Sequence 5, Appli
33	42	56.0	6	1	US-08-188-223-4 Sequence 4, Appli
34	42	56.0	6	4	US-08-968-466-4 Sequence 4, Appli
35	42	56.0	6	4	US-08-478-546B-4 Sequence 4, Appli
36	42	56.0	6	5	PCT-US94-13205-5 Sequence 5, Appli
37	42	56.0	135	4	US-09-281-766-10 Sequence 10, Appli
38	42	56.0	320	6	5223391-3 Patent No. 5223391
39	42	56.0	865	4	US-09-281-766-19 Sequence 19, Appli
40	42	56.0	1235	2	US-08-680-326-36 Sequence 36, Appli
41	41.5	55.3	21	4	US-08-602-999A-427 Sequence 427, App
42	41.5	55.3	21	4	US-09-500-124-427 Sequence 427, App
43	41	54.7	54	3	US-08-301-162-12 Sequence 12, Appli
44	41	54.7	54	4	US-09-461-240-12 Sequence 12, Appli
45	41	54.7	54	4	US-09-968-927-12 Sequence 12, Appli

ALIGNMENTS

RESULT 1  
US-08-446-692-73  
; Sequence 73, Application US/08446692  
; Patent No. 5759551

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Y1

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maria C.H. Lin

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/446,692

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4146 US2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)415-8745

TELEFAX: (516)751-6849

INFORMATION FOR SEQ ID NO: 73:

SEQUENCE CHARACTERISTICS:

LENGTH: 11 amino acids

TYPE: amino acid

TOPOLOGY: Linear

MOLECULE TYPE: peptide

US-08-446-692-73

Query Match 74.7%; Score 56; DB 1; Length 11;  
Best Local Similarity 81.8%; Pred. No. 0.077;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ELGPGGRPPPP 11  
Db 1 QLGPGGRPPPP 11

RESULT 2  
US-08-488-351A-73  
Sequence 73, Application US/08488351A  
Patent No. 5843446  
GENERAL INFORMATION:  
APPLICANT: Ladd, Anna I  
APPLICANT: Wang, Chang Yi  
APPLICANT: Zamb, Timothy  
TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Maria C.H. Lin  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10154-0053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488.351A  
FILING DATE: 7-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/446.692  
FILING DATE: 7-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/229.275  
FILING DATE: 14-APR-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/057.166  
FILING DATE: 27-APR-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maria C.H. Lin  
REGISTRATION NUMBER: 29.323  
REFERENCE/DOCKET NUMBER: 1151-4146 US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)415-8745  
TELEFAX: (516)751-6849  
INFORMATION FOR SEQ ID NO: 73:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-488-351A-73

Query Match 74.7%; Score 56; DB 2; Length 11;  
Best Local Similarity 81.8%; Pred. No. 0.077;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ELGPGGPPPP 11  
:|||||  
Db 1 QLGPGGPPPP 11

RESULT 3  
US-08-188-223-2  
Sequence 2, Application US/08188223  
Patent No. 5688506  
GENERAL INFORMATION:  
APPLICANT: Grimes, Stephen  
APPLICANT: Scibienski, Robert  
TITLE OF INVENTION: Immunogens Against Gonadotropin  
TITLE OF INVENTION: Releasing Hormone

NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dimitrios T. Drivas, Esq.  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/188.223  
FILING DATE: 27-JAN-1994  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Drivas Esq., Dimitrios T.  
REGISTRATION NUMBER: 32.218  
REFERENCE/DOCKET NUMBER: 1102865-300  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-819-8286  
TELEFAX: 212-354-8113  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: YES  
FEATURE:  
NAME/KEY: Region  
LOCATION: 1..10  
OTHER INFORMATION: /note= "immunomimic"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 11..16  
OTHER INFORMATION: /note= "spacer"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label= pglu  
US-08-188-223-2

Query Match 64.0%; Score 48; DB 1; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GRPPPC 12  
:|||||  
Db 10 GRPPPC 16

RESULT 4  
US-08-968-466-2  
Sequence 2, Application US/08968466  
Patent No. 6132720  
GENERAL INFORMATION:  
APPLICANT: Grimes, Stephen  
APPLICANT: Scibienski, Robert  
TITLE OF INVENTION: Immunogens Against Gonadotropin  
TITLE OF INVENTION: Releasing Hormone  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dimitrios T. Drivas, Esq.  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2787  
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: 1  
APPLICATION NUMBER: US/08/968,466  
FILING DATE: 27-JAN-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Drivas Esq., Dimitrios T.  
REGISTRATION NUMBER: 32,218  
REFERENCE/DOCKET NUMBER: 1102865-300  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-354-8113  
TELEFAX: 212-354-8286  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: YES  
FEATURE:  
NAME/KEY: Region  
LOCATION: 1..10  
OTHER INFORMATION: /note= "immunomimic"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 11..16  
OTHER INFORMATION: /note= "spacer"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label= pglu  
OTHER INFORMATION: /note= "pyroglutamic acid (5-oxoproline)"  
US-08-968-466-2  
Query Match 64.0%; Score 48; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GRPPPC 12  
DB 10 GRPPPC 16  
RESULT 5  
US-08-478-546B-2  
Sequence 2, Application US/08478546B  
Patent No. 6303123  
GENERAL INFORMATION:  
APPLICANT: Grimes, Stephen  
TITLE OF INVENTION: Methods for the Treatment of Hormone-Dependent  
TITLE OF INVENTION: Tumors with Immunogens against Gonadotropin Releasing Hormone  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dimitrios T. Drivas, Esq.  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,546B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/188,223  
FILING DATE: 27-JAN-1994  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Drivas Esq., Dimitrios T.  
REGISTRATION NUMBER: 32,218  
REFERENCE/DOCKET NUMBER: 1102865-300  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-354-8113  
TELEFAX: 212-354-8286  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified site  
LOCATION: 1  
OTHER INFORMATION: /note= Xaa  
OTHER INFORMATION: /note= "pyroglutamic acid"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 1..10  
OTHER INFORMATION: /note= "immunomimic"  
FEATURE:  
NAME/KEY: Region  
LOCATION: 11..16  
OTHER INFORMATION: /note= "spacer"  
US-08-478-546B-2  
Query Match 64.0%; Score 48; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 6 GRPPPC 12  
DB 10 GRPPPC 16  
RESULT 6  
US-07-832-855-2  
Sequence 2, Application US/07832855  
Patent No. 5478727  
GENERAL INFORMATION:  
APPLICANT: Roizman, Bernard  
APPLICANT: Liu, Fenyong  
TITLE OF INVENTION: Methods and Compositions of a  
TITLE OF INVENTION: Preparation and Use of A Herpes Protease  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE & DURKEE  
STREET: 321 No. 5478727th Clark Street, Suite 800  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/832,855  
FILING DATE: 19920207  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Cooley, Ronald B.  
REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD045  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4961

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 635 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-832-855-2

Query Match  
Best Local Similarity 62.7%; Score 47; DB 1; Length 635;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 POGRRPPPC 12  
| | | | |  
Db 565 POGRRPPPC 573

RESULT 7  
US-08-176-320-2  
Sequence 2, Application US/08176320  
Patent No. 6410704  
GENERAL INFORMATION:  
APPLICANT: Roizman, Bernard  
APPLICANT: Liu, Fengyong  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
NUMBER OF SEQUENCES: 2  
PREPARATION AND USE OF A HERPES PROTEASE  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Alice O. Martin  
STREET: 321 No. 6410704th Clark Street, Suite 800  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60610

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/176,320  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/705,814  
FILING DATE: 24-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Cooley, Ronald B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312)744-0090  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 635 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-176-320-2

Query Match  
Best Local Similarity 62.7%; Score 47; DB 4; Length 635;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 POGRRPPPC 12  
| | | | |  
Db 565 POGRRPPPC 573

RESULT 8  
US-08-457-273B-8  
Sequence 8, Application US/08457273B  
Patent No. 5849995  
GENERAL INFORMATION:  
APPLICANT: Hayden, Michael

APPLICANT: Lin, Biaoyang  
APPLICANT: Nasir, Jamal  
TITLE OF INVENTION: Mouse Model for Huntington's Disease and  
TITLE OF INVENTION: Related DNA Sequences  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Virginia Bennett  
STREET: PO Box 37428  
CITY: Raleigh  
STATE: No. 5849995th Carolina  
COUNTRY: US  
ZIP: 27627  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,273B  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Bennett, Virginia C.  
REGISTRATION NUMBER: 37,092  
REFERENCE/DOCKET NUMBER: 3477-85A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-854-1400  
TELEFAX: 919-854-1401  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 318 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-457-273B-8

Query Match  
Best Local Similarity 62.7%; Score 47; DB 2; Length 318;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 POGRRPPPC 11  
| | | | |  
Db 45 POGRRPPPC 52

RESULT 9  
US-08-246-982A-16  
Sequence 16, Application US/08246982A  
Patent No. 5686288  
GENERAL INFORMATION:  
APPLICANT: MacDonald, Marcy E.  
APPLICANT: Ambrose, Christine M.  
APPLICANT: Duyao, Mabel P.  
APPLICANT: Gusella, James F.  
TITLE OF INVENTION: Huntinglin DNA, Protein And Uses Thereof  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/246,982A  
FILING DATE: May 20, 1994  
CLASSIFICATION: 435

```

; ATTORNEY/AGENT INFORMATION:
;   NAME: Goldstein, Jorge, A.
;   REGISTRATION NUMBER: 29,021
;   REFERENCE/DOCKET NUMBER: 0609.3880002
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600
;   TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 3119 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-246-982A-16

Query Match
Best Local Similarity 62.7%; Score 47; DB 1; Length 3119;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 POGRRPPP 11
Db 45 POGRRPPP 52

RESULT 10
US-08-453-265-16
; Sequence 16, Application US/08453265
; Patent No. 5693757
; GENERAL INFORMATION:
;   APPLICANT: MacDonald, Marcy E.
;   APPLICANT: Ambrose, Christine M.
;   APPLICANT: Duyao, Mabel P.
;   APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Huntington DNA, Protein And Uses Thereof
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,265
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
;   NAME: Ludwig, Steven R.
;   REGISTRATION NUMBER: 36,203
;   REFERENCE/DOCKET NUMBER: 0609.3880003
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600
;   TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 3119 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-453-265-16

Query Match
Best Local Similarity 62.7%; Score 47; DB 1; Length 3119;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 POGRRPPP 11
Db 45 POGRRPPP 52
```

```

RESULT 11
US-08-237-919-2
; Sequence 2, Application US/08237919
; Patent No. 5610281
; GENERAL INFORMATION:
;   APPLICANT: Brenner, Michael B
;   APPLICANT: Ceppek, Karyn L
; TITLE OF INVENTION: Methods and Compositions for
;   Modulating Heterotypic E-cadherin Interactions with T Lymph
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

Query Match
Best Local Similarity 61.3%; Score 46; DB 1; Length 878;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 LGPQGRPPPC 12
Db 19 LGQGRSPPC 29

RESULT 12
US-08-732-429-2
; Sequence 2, Application US/08732429
; Patent No. 6300080
; GENERAL INFORMATION:
;   APPLICANT:
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
;   E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
```

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/732,429  
;; FILING DATE: herewith  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/237,919  
;; FILING DATE: 3 May 1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Plumer, Elizabeth R  
;; REGISTRATION NUMBER: 36,637  
;; REFERENCE/DOCKET NUMBER: B0801/7023  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617-720-3500  
;; TELEFAX: 617-720-2441  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 878 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-732-429-2

Query Match 61.3%; Score 46; DB 4; Length 878;  
Best Local Similarity 72.7%; Pred. No. 99;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 LGPGRRPPPC 12  
Db 19 LGSQERSPPC 29

RESULT 13  
US-09-798-267-2  
;; Sequence 2, Application US/09798267  
;; Patent No. 6406870  
;; GENERAL INFORMATION:  
;; APPLICANT: Brenner, Michael  
;; APPLICANT: Cepek, Karyn  
;; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin  
;; FILE REFERENCE: L0560/7008ERP  
;; CURRENT APPLICATION NUMBER: US/09/798,267  
;; CURRENT FILING DATE: 2001-03-02  
;; PRIOR APPLICATION NUMBER: US 08/237,919  
;; PRIOR FILING DATE: 1994-05-03  
;; PRIOR APPLICATION NUMBER: PCT/US 95/05518  
;; PRIOR FILING DATE: 1995-05-03  
;; PRIOR APPLICATION NUMBER: US 08/732,429  
;; PRIOR FILING DATE: 1996-11-01  
;; NUMBER OF SEQ ID NOS: 9  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 2  
;; LENGTH: 878  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-798-267-2

Query Match 61.3%; Score 46; DB 4; Length 878;  
Best Local Similarity 72.7%; Pred. No. 99;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 LGPGRRPPPC 12  
Db 19 LGSQERSPPC 29

RESULT 14  
US-09-798-267-3  
;; Sequence 3, Application US/09798267  
;; Patent No. 6406870  
;; GENERAL INFORMATION:  
;; APPLICANT: Brenner, Michael  
;; APPLICANT: Cepek, Karyn  
;; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin

;; TITLE OF INVENTION: Interactions with T Lymphocytes  
;; FILE REFERENCE: L0560/7008ERP  
;; CURRENT APPLICATION NUMBER: US/09/798,267  
;; CURRENT FILING DATE: 2001-03-02  
;; PRIOR APPLICATION NUMBER: US 08/237,919  
;; PRIOR FILING DATE: 1994-05-03  
;; PRIOR APPLICATION NUMBER: PCT/US 95/05518  
;; PRIOR FILING DATE: 1995-05-03  
;; PRIOR APPLICATION NUMBER: US 08/732,429  
;; PRIOR FILING DATE: 1996-11-01  
;; NUMBER OF SEQ ID NOS: 9  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 3  
;; LENGTH: 878  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-798-267-3

Query Match 61.3%; Score 46; DB 4; Length 878;  
Best Local Similarity 72.7%; Pred. No. 99;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 LGPGRRPPPC 12  
Db 19 LGSQERSPPC 29

RESULT 15  
PCT-US95-05518-2  
;; Sequence 2, Application PC/TUS9505518  
;; GENERAL INFORMATION:  
;; APPLICANT:  
;; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic  
;; NUMBER OF INVENTION: E-cadherin Interactions with T Lymphocytes  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Wolf, Greenfield & Sacks P.C.  
;; STREET: 600 Atlantic Avenue  
;; CITY: Boston  
;; STATE: MA  
;; COUNTRY: USA  
;; ZIP: 02210  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US95/05518  
;; FILING DATE: herewith  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/237,919  
;; FILING DATE: 3 May 1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Plumer, Elizabeth R  
;; REGISTRATION NUMBER: 36,637  
;; REFERENCE/DOCKET NUMBER: B0801/7023  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617-720-3500  
;; TELEFAX: 617-720-2441  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 878 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
PCT-US95-05518-2

Query Match 61.3%; Score 46; DB 5; Length 878;  
Best Local Similarity 72.7%; Pred. No. 99;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 LGPGRRPPPC 12

|||  
Db 19 LSGQERSPPC 29

Search completed: February 5, 2003, 09:50:19  
Job time : 15.1429 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:49:21 : Search time 9.14286 Seconds

(without alignments)  
29.097 Million cell updates/sec

Title: US-09-700-329-2

Perfect score: 75

Sequence: 1 ELGPGRRPPPC 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 129505 seqs, 22169297 residues

Total number of hits satisfying chosen parameters: 129505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PC1\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/PC1US\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	65.3	11	9	US-10-192-257-2
2	46	61.3	878	9	US-10-165-049-2
3	46	61.3	878	10	US-10-165-049-3
4	46	61.3	878	10	US-09-905-983-48
5	45	60.0	638	9	US-10-001-887-108
6	45	60.0	1806	10	US-09-919-497-56
7	44	58.7	34	10	US-09-905-831-21
8	44	58.7	56	9	US-10-001-887-137
9	44	58.7	322	10	US-09-823-038A-33
10	44	58.7	1249	9	US-09-964-889-33
11	43	57.3	274	10	US-09-850-887-4
12	43	57.3	1142	10	US-09-894-998-3
13	42	56.0	6	9	US-10-192-257-5
14	42	56.0	9	9	US-10-192-257-3
15	42	56.0	12	9	US-10-192-257-1
16	42	56.0	135	9	US-09-957-995A-10
17	42	56.0	209	10	US-09-764-864-800
18	42	56.0	292	10	US-09-764-864-1261
19	42	56.0	389	10	US-09-788-657-23

20	42	56.0	531	10	US-09-925-300-1444	Sequence 1444, Ap
21	42	56.0	802	9	US-09-823-240-2	Sequence 2, Appli
22	42	56.0	865	9	US-09-957-995A-19	Sequence 19, Appli
23	42	56.0	1235	10	US-09-904-065-6	Sequence 6, Appli
24	42	56.0	1235	10	US-09-904-065-8	Sequence 8, Appli
25	42	56.0	1235	10	US-09-904-065-10	Sequence 10, Appli
26	42	56.0	1235	10	US-09-904-065-16	Sequence 16, Appli
27	42	56.0	1235	10	US-09-904-065-17	Sequence 17, Appli
28	42	56.0	1235	10	US-09-904-065-18	Sequence 18, Appli
29	42	56.0	1235	10	US-09-904-065-19	Sequence 19, Appli
30	42	56.0	1884	10	US-09-785-770A-17	Sequence 17, Appli
31	42	56.0	1907	10	US-09-785-770A-16	Sequence 16, Appli
32	42	56.0	4019	9	US-09-854-133-425	Sequence 425, App
33	42	56.0	4019	10	US-09-738-973-425	Sequence 425, App
34	41.5	55.3	173	10	US-09-925-300-962	Sequence 962, App
35	41.5	55.3	349	10	US-09-80-996-7	Sequence 6, Appli
36	41	54.7	52	10	US-09-864-761-43362	Sequence 43362, A
37	41	54.7	111	10	US-09-867-550-1410	Sequence 1410, Ap
38	41	54.7	125	10	US-09-764-887-279	Sequence 279, App
39	41	54.7	171	10	US-09-925-301-1078	Sequence 1078, Ap
40	41	54.7	217	10	US-09-770-906-2	Sequence 2, Appli
41	41	54.7	278	10	US-09-246-129B-6	Sequence 6, Appli
42	41	54.7	278	10	US-09-899-059-6	Sequence 6, Appli
43	41	54.7	279	8	US-08-971-317A-4	Sequence 4, Appli
44	41	54.7	279	9	US-10-017-910-5	Sequence 5, Appli
45	41	54.7	279	10	US-09-193-663-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-10-192-257-2

Sequence 2, Application US/10192257

Publication No. US20030021786A1

GENERAL INFORMATION:

APPLICANT: Aphion Corporation

TITLE OF INVENTION: Treatment and Prevention of Cancerous and Pre-Cancerous Condit

FILE REFERENCE: 1102865-0057

CURRENT APPLICATION NUMBER: US/10/192,257

PRIOR FILING DATE: 2002-07-09

PRIOR APPLICATION NUMBER: US 60/303,868

NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn version 3.1

SEQ ID NO 2

LENGTH: 11

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: MISC\_FEATURE

LOCATION: (1)..(1)

OTHER INFORMATION: Xaa=pyroglutamic acid

US-10-192-257-2

Query Match

Best Local Similarity

Matches 8; Conservative

Db 2 GPGRRPPPC 12

US-10-165-049-2

Sequence 2, Application US/10165049

Publication No. US20020192724A1

GENERAL INFORMATION:

APPLICANT: Brenner, Michael

TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin

;; TITLE OF INVENTION: Interactions with T Lymphocytes  
;; FILE REFERENCE: L00560/70010ERP  
;; CURRENT APPLICATION NUMBER: US/10/165,049  
;; PRIOR FILING DATE: 2002-06-07  
;; PRIOR APPLICATION NUMBER: USSN 08/237,919  
;; PRIOR FILING DATE: 1994-05-03  
;; PRIOR APPLICATION NUMBER: PCT/US 95/05518  
;; PRIOR FILING DATE: 1995-05-03  
;; PRIOR APPLICATION NUMBER: USSN 08/732,429  
;; PRIOR FILING DATE: 1996-11-01  
;; PRIOR APPLICATION NUMBER: USSN 09/798,267  
;; PRIOR FILING DATE: 2001-03-01  
;; NUMBER OF SEQ ID NOS: 9  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 2  
;; LENGTH: 878  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (800)..(808)  
;; OTHER INFORMATION: HAV tripeptide  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (2225)..(2295)  
;; OTHER INFORMATION: transmembrane  
US-10-165-049-2

Query Match 61.3%; Score 46; DB 9; Length 878;  
Best Local Similarity 72.7%; Pred. No. 96;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 LGPQGRPPPC 12  
||| | |||||  
DB 19 LGSQERSPPC 29

RESULT 3  
US-10-165-049-3  
;; Sequence 3, Application US/10165049  
;; Publication No. US20020192724A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Brenner, Michael  
;; APPLICANT: Cepek, Karyn  
;; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin  
;; FILE REFERENCE: L00560/70010ERP  
;; CURRENT APPLICATION NUMBER: US/10/165,049  
;; PRIOR FILING DATE: 2002-06-07  
;; PRIOR APPLICATION NUMBER: USSN 08/237,919  
;; PRIOR FILING DATE: 1994-05-03  
;; PRIOR APPLICATION NUMBER: PCT/US 95/05518  
;; PRIOR FILING DATE: 1995-05-03  
;; PRIOR APPLICATION NUMBER: USSN 08/732,429  
;; PRIOR FILING DATE: 1996-11-01  
;; PRIOR APPLICATION NUMBER: USSN 09/798,267  
;; PRIOR FILING DATE: 2001-03-01  
;; NUMBER OF SEQ ID NOS: 9  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 3  
;; LENGTH: 878  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-165-049-3

Query Match 61.3%; Score 46; DB 9; Length 878;  
Best Local Similarity 72.7%; Pred. No. 96;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 LGPQGRPPPC 12  
||| | |||||  
DB 19 LGSQERSPPC 29

RESULT 4  
US-09-905-983-48  
;; Sequence 48, Application US/09905983  
;; Patent No. US20020045591A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Geiger, Benjamin  
;; APPLICANT: Ben-Ze'ev, Avri  
;; APPLICANT: Sadot, Eilat  
;; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER  
;; FILE REFERENCE: 01/22326  
;; CURRENT APPLICATION NUMBER: US/09/905,983  
;; PRIOR FILING DATE: 2001-09-28  
;; NUMBER OF SEQ ID NOS: 60  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 48  
;; LENGTH: 878  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-905-983-48

Query Match 61.3%; Score 46; DB 10; Length 878;  
Best Local Similarity 72.7%; Pred. No. 96;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 LGPQGRPPPC 12  
||| | |||||  
DB 19 LGSQERSPPC 29

RESULT 5  
US-10-001-887-108  
;; Sequence 108, Application US/10001887  
;; Patent No. US20020155464A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Salceda, Susana  
;; APPLICANT: Macina, Roberto  
;; APPLICANT: Recipon, Herve  
;; APPLICANT: Cafierkey, Robert  
;; APPLICANT: Sun, Yongming  
;; APPLICANT: Liu, Chenghua  
;; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and  
;; FILE REFERENCE: DEX-0269  
;; CURRENT APPLICATION NUMBER: US/10/001,887  
;; PRIOR FILING DATE: 2001-11-20  
;; PRIOR APPLICATION NUMBER: 60/249,998  
;; PRIOR FILING DATE: 2000-11-20  
;; PRIOR APPLICATION NUMBER: 60/252,563  
;; PRIOR FILING DATE: 2000-11-22  
;; NUMBER OF SEQ ID NOS: 137  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 108  
;; LENGTH: 638  
;; TYPE: PRT  
;; ORGANISM: Homo sapien  
US-10-001-887-108

Query Match 60.0%; Score 45; DB 9; Length 638;  
Best Local Similarity 72.7%; Pred. No. 95;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ELGPQGRPPP 11  
||| | |||  
DB 124 EPQPGERPPP 134

RESULT 6  
US-09-919-497-56  
;; Sequence 56, Application US/09919497  
;; Patent No. US20020106662A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Mutter, George L.  
;; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER

```
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (758)..(758)
; OTHER INFORMATION: Xaa = any amino acid
; NAME/KEY: UNSURE
; LOCATION: (809)..(809)
; OTHER INFORMATION: Xaa = any amino acid
US-09-919-497-56
```

```
Query Match          60.0%; Score 45; DB 10; Length 1806;
Best Local Similarity 70.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 2 LGPGRRPPPP 11
    ||| |||
Db 681 MGPGRRPPPP 690
```

```
RESULT 7
US-09-905-831-21
; Sequence 21, Application US/09905831
; Patent No. US20020119572A1
; GENERAL INFORMATION:
; APPLICANT: Jacobson, Joseph
; APPLICANT: Schwartz, John
; APPLICANT: Hamad, Kimberly
; APPLICANT: Zhang, Shuangang
; TITLE OF INVENTION: Direct, Externally Imposed Control of Polypeptides
; FILE REFERENCE: MLB-086
; CURRENT APPLICATION NUMBER: US/09/905,831
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/276,313
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/218,312
; PRIOR FILING DATE: 2000-07-14
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PXY34 peptide
US-09-905-831-21
```

```
Query Match          58.7%; Score 44; DB 10; Length 34;
Best Local Similarity 70.0%; Pred. No. 8.5;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY 3 GPGRRPPPPC 12
    || | |||
Db 25 GPPGPPPPPC 34
```

```
RESULT 8
US-10-001-887-137
; Sequence 137, Application US/10001887
; Patent No. US20020155464A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herive
```

```
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and
; FILE REFERENCE: DEX-0269
; CURRENT APPLICATION NUMBER: US/10/001,887
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,998
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,563
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 137
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-887-137
```

```
Query Match          58.7%; Score 44; DB 9; Length 56;
Best Local Similarity 77.8%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 3 GPGRRPPPP 11
    || | |||
Db 8 GPGRRPPPP 16
```

```
RESULT 9
US-09-823-038A-33
; Sequence 33, Application US/09823038A
; Patent No. US2002005835A1
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Abernethy, Nevin
; APPLICANT: Oprust, Rene
; APPLICANT: Kumble, Anand
; APPLICANT: Morrison, Greg
; TITLE OF INVENTION: Compositions Isolated From Stromal Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1037c3
; CURRENT APPLICATION NUMBER: US/09/823,038A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Human
US-09-823-038A-33
```

```
Query Match          58.7%; Score 44; DB 10; Length 322;
Best Local Similarity 77.8%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 3 GPGRRPPPP 11
    || | |||
Db 16 GPGRRPPPP 24
```

```
RESULT 10
US-09-964-899-33
; Sequence 33, Application US/09964899
; Patent No. US2002017446A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Dalia et al.
; TITLE OF INVENTION: Identification of Genes Involved in
; FILE REFERENCE: 4-31612 A
; CURRENT APPLICATION NUMBER: US/09/964,899
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/236,893
```

PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/298,309  
PRIOR FILING DATE: 2001-06-14  
NUMBER OF SEQ ID NOS: 53  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 33  
LENGTH: 1249  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-09-964-899-33

Query Match 58.7%; Score 44; DB 9; Length 1249;  
Best Local Similarity 63.6%; Pred. No. 2.3e+02;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 ELPGGRPPPP 11  
11:111111  
DB 839 EIPRRRPPPP 849

## RESULT 11

US-09-850-887-4  
Sequence 4, Application US/09850887  
Patent No. US20020009778A1  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti

Corley, Neil C.

TITLE OF INVENTION: THROID AND PITUITARY MEMBRANE PROTEIN  
NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 PORTER DRIVE

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/850,887

FILING DATE: 07-May-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/087,678

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: CERRONE, MICHAEL C.

REGISTRATION NUMBER: 39,132

REFERENCE/DOCKET NUMBER: PP-0535 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 274 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: g206712

SEQUENCE DESCRIPTION: SEQ ID NO: 4 :

US-09-850-887-4

Query Match 57.3%; Score 43; DB 10; Length 274;  
Best Local Similarity 87.5%; Pred. No. 77;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GPGGRPPP 10

DB 192 GPGGRPPP 199

## RESULT 12

US-09-894-998-3  
Sequence 3, Application US/09894998  
Patent No. US20020090610A1  
GENERAL INFORMATION:

APPLICANT: Hosken, Nancy Ann

APPLICANT: Craig H. Day

APPLICANT: David C. Dillon

APPLICANT: McGowan, Patrick

APPLICANT: Sleath, Paul R.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND

FILE REFERENCE: 210121.538

CURRENT APPLICATION NUMBER: US/09/894,998

CURRENT FILING DATE: 2001-06-28

NUMBER OF SEQ ID NOS: 64

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 1142

TYPE: PRT

ORGANISM: Herpes simplex virus

US-09-894-998-3

Query Match 57.3%; Score 43; DB 10; Length 1142;  
Best Local Similarity 77.8%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 GPGGRPPP 11  
11111111  
DB 146 GPGGRPPP 154

## RESULT 13

US-10-192-257-5  
Sequence 5, Application US/10192257  
Publication No. US20030021786A1  
GENERAL INFORMATION:

APPLICANT: Aption Corporation

TITLE OF INVENTION: Treatment and Prevention of Cancerous and Pre-Cancerous Condit

FILE REFERENCE: 1102865-0057

CURRENT APPLICATION NUMBER: US/10/192,257

CURRENT FILING DATE: 2002-07-09

PRIOR APPLICATION NUMBER: US 60/303,868

PRIOR FILING DATE: 2001-07-09

NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn version 3.1

SEQ ID NO 5

LENGTH: 6

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Hypothetical spacer peptide

US-10-192-257-5

Query Match 56.0%; Score 42; DB 9; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.1e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 RPPPPC 12  
11111111  
DB 1 RPPPPC 6

## RESULT 14

US-10-192-257-3  
Sequence 3, Application US/10192257  
Publication No. US20030021786A1  
GENERAL INFORMATION:

```
; APPLICANT: Aphion Corporation
; TITLE OF INVENTION: Treatment and Prevention of Cancerous and Pre-Cancerous Condition
; FILE REFERENCE: 1102865-0057
; CURRENT APPLICATION NUMBER: US/10/192,257
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: US 60/303,868
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa=pyroglutamic acid
US-10-192-257-3
```

```
Query Match          56.0%; Score 42; DB 9; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 7 RPPPPC 12
    |||||
Db  4 RPPPPC 9
```

```
RESULT 15
US-10-192-257-1
; Sequence 1, Application US/10192257
; Publication No. US20030021786A1
; GENERAL INFORMATION:
; APPLICANT: Aphion Corporation
; TITLE OF INVENTION: Treatment and Prevention of Cancerous and Pre-Cancerous Condition
; FILE REFERENCE: 1102865-0057
; CURRENT APPLICATION NUMBER: US/10/192,257
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: US 60/303,868
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa=pyroglutamic acid
US-10-192-257-1
```

```
Query Match          56.0%; Score 42; DB 9; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 7 RPPPPC 12
    |||||
Db  7 RPPPPC 12
```

```
Search completed: February 5, 2003, 09:56:02
Job time : 10.1429 secs
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model1

Run on: February 5, 2003, 09:48:26 : Search time 156 Seconds  
(Without alignments)  
49.595 Million cell updates/sec

Title: US-09-700-329-2

Perfect score: 75

Sequence: 1 ELGPGGRPPPC 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

```
1: /cgn2_6/ptodata/1/paa/PCrUS.COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06.COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07.COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US08.COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US081.COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US082.COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US083.COMB.pep:*
8: /cgn2_6/ptodata/1/paa/US084.COMB.pep:*
9: /cgn2_6/ptodata/1/paa/US085.COMB.pep:*
10: /cgn2_6/ptodata/1/paa/US086.COMB.pep:*
11: /cgn2_6/ptodata/1/paa/US087.COMB.pep:*
12: /cgn2_6/ptodata/1/paa/US088.COMB.pep:*
13: /cgn2_6/ptodata/1/paa/US089.COMB.pep:*
14: /cgn2_6/ptodata/1/paa/US090.COMB.pep:*
15: /cgn2_6/ptodata/1/paa/US091.COMB.pep:*
16: /cgn2_6/ptodata/1/paa/US092.COMB.pep:*
17: /cgn2_6/ptodata/1/paa/US093.COMB.pep:*
18: /cgn2_6/ptodata/1/paa/US094.COMB.pep:*
19: /cgn2_6/ptodata/1/paa/US095.COMB.pep:*
20: /cgn2_6/ptodata/1/paa/US096.COMB.pep:*
21: /cgn2_6/ptodata/1/paa/US097.COMB.pep:*
22: /cgn2_6/ptodata/1/paa/US098.COMB.pep:*
23: /cgn2_6/ptodata/1/paa/US099.COMB.pep:*
24: /cgn2_6/ptodata/1/paa/US100.COMB.pep:*
25: /cgn2_6/ptodata/1/paa/US101.COMB.pep:*
26: /cgn2_6/ptodata/1/paa/US102.COMB.pep:*
27: /cgn2_6/ptodata/1/paa/US60.COMB.pep:*
```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	100.0	12	1	PCT-US99-10734-5
2	75	100.0	12	1	PCT-US99-10751-2
3	75	100.0	12	21	US-09-700-329-2
4	75	100.0	12	21	US-09-700-378-5
5	56	74.7	11	1	PCT-US94-04832A-73
6	56	74.7	11	8	US-08-488-320A-73

7	56	74.7	160	1	PCT-US01-14827-14717	Sequence 14717, A
8	55	73.3	139	1	PCT-US01-08656-6637	Sequence 6637, Ap
9	55	73.3	169	1	PCT-US01-08656-6639	Sequence 6639, Ap
10	52	69.3	185	21	US-09-758-444-470	Sequence 470, App
11	52	69.3	323	21	US-09-758-444-476	Sequence 476, App
12	52	69.3	323	26	US-10-212-160-476	Sequence 476, App
13	51.5	68.7	172	24	US-10-015-127-10407	Sequence 10407, A
14	51	68.0	144	1	PCT-US99-22855-962	Sequence 962, App
15	51	68.0	222	18	US-09-451-390-4290	Sequence 4290, Ap
16	51	68.0	250	1	PCT-US99-22855-961	Sequence 961, App
17	51	68.0	261	21	US-09-791-357-140536	Sequence 140536, Ap
18	51	68.0	261	27	US-60-356-051-1558	Sequence 1558, Ap
19	50	66.7	163	23	US-09-924-340-58	Sequence 58, Appl
20	50	66.7	163	23	US-09-992-600A-58	Sequence 58, Appl
21	50	66.7	163	23	US-09-994-590-58	Sequence 58, Appl
22	50	66.7	163	24	US-10-000-489-58	Sequence 58, Appl
23	50	66.7	163	24	US-10-000-966-58	Sequence 58, Appl
24	50	66.7	163	27	US-60-305-456-58	Sequence 58, Appl
25	50	66.7	163	27	US-60-212-655-594	Sequence 594, App
26	49	65.3	11	1	PCT-US02-21768-2	Sequence 2, Appli
27	49	65.3	11	1	PCT-US99-10734-2	Sequence 2, Appli
28	49	65.3	11	11	US-08-798-423-4	Sequence 4, Appli
29	49	65.3	11	21	US-09-700-329-6	Sequence 6, Appli
30	49	65.3	11	21	US-09-700-378-2	Sequence 2, Appli
31	49	65.3	11	25	US-10-192-257-2	Sequence 2, Appli
32	49	65.3	383	1	PCT-US00-08979-208	Sequence 208, App
33	49	65.3	383	1	PCT-US00-35017A-1128	Sequence 1128, Ap
34	49	65.3	383	20	US-09-684-524-239	Sequence 239, App
35	49	65.3	383	24	US-10-050-704-239	Sequence 239, App
36	49	65.3	1233	1	PCT-US02-25485-2856	Sequence 2856, App
37	49	65.3	1233	1	PCT-US02-30474-3124	Sequence 3124, Ap
38	49	65.3	1233	27	US-60-311-261-2719	Sequence 2719, Ap
39	48	64.0	16	1	PCT-US95-01225-2	Sequence 2, Appli
40	48	64.0	16	17	US-09-305-924-5	Sequence 5, Appli
41	48	64.0	16	17	US-09-306-688-5	Sequence 5, Appli
42	48	64.0	72	1	PCT-US01-42950-593	Sequence 593, App
43	48	64.0	112	16	US-09-270-767-31687	Sequence 31687, A
44	48	64.0	112	16	US-09-270-767-46904	Sequence 46904, A
45	48	64.0	112	16	US-09-270-849B-180163	Sequence 180163, A

#### ALIGNMENTS

```
RESULT 1
PCT-US99-10734-5
: Sequence 5, Application PC/TUS9910734
:
: GENERAL INFORMATION:
: APPLICANT: Gevas, Phillip C.
: APPLICANT: Grimes, Stephen
: APPLICANT: Karr, Stephen
: APPLICANT: Michaeli, Dov
: TITLE OF INVENTION: Method for the Treatment of Gastroesophageal Reflux
: FILE REFERENCE: ACGI/PC
: CURRENT APPLICATION NUMBER: PCT/US99/10734
: EARLIER APPLICATION NUMBER: 60/085,610
: EARLIER FILING DATE: 1998-05-15
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 5
: LENGTH: 12
: TYPE: PRT
: ORGANISM: Unknown
: FEATURE:
: NAME/KEY: MOD_RES
: LOCATION: (1)
: OTHER INFORMATION: pyroglutamic acid
: FEATURE:
: OTHER INFORMATION: Description of Unknown Organism: human or
: OTHER INFORMATION: synthetic peptide
```

PCT-US99-10734-5

Query Match 100.0%; Score 75; DB 1; Length 12;

Best Local Similarity 100.0%; Pred. No. 0.04;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ELGPGGRPPPC 12

Db 1 ELGPGGRPPPC 12

RESULT 2

PCT-US99-10751-2

Sequence 2, Application PC/TUS9910751

GENERAL INFORMATION:

APPLICANT: APHTON CORPORATION

TITLE OF INVENTION: Prevention and Treatment of Hypergastrinemia

FILE REFERENCE: 1102865-0035

CURRENT APPLICATION NUMBER: PCT/US99/10751

CURRENT FILING DATE: 1999-05-14

EARLIER APPLICATION NUMBER: US 60/085,714

PRIOR FILING DATE: 1999-05-15

NUMBER OF SEQ ID NOS: 4

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 12

TYPE: PRT

ORGANISM: human or synthetic peptide

FEATURE:

NAME/KEY: MOD\_RES

LOCATION: (1)

OTHER INFORMATION: pyroglutamic acid

PCT-US99-10751-2

Query Match 100.0%; Score 75; DB 1; Length 12;

Best Local Similarity 100.0%; Pred. No. 0.04;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ELGPGGRPPPC 12

Db 1 ELGPGGRPPPC 12

RESULT 3

US-09-700-329-2

Sequence 2, Application US/09700329

GENERAL INFORMATION:

APPLICANT: Gevas, Philip

APPLICANT: Grimes, Stephen

APPLICANT: Karr, Stephen

APPLICANT: Michaeli, Dov

APPLICANT: Watson, Susan

TITLE OF INVENTION: Prevention and Treatment of Hypergastrinemia

FILE REFERENCE: ACG2USA

CURRENT APPLICATION NUMBER: US/09/700,329

CURRENT FILING DATE: 2001-02-08

PRIOR APPLICATION NUMBER: PCT/US99/10751

PRIOR FILING DATE: 1999-05-14

PRIOR APPLICATION NUMBER: 60/085,714

PRIOR FILING DATE: 1998-05-15

NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn version 3.0

SEQ ID NO 2

LENGTH: 12

TYPE: PRT

ORGANISM: human gastrin peptide

FEATURE:

NAME/KEY: MOD\_RES

LOCATION: (1)-(1)

OTHER INFORMATION: pyroglutamic acid

US-09-700-329-2

Query Match 100.0%; Score 75; DB 21; Length 12;

Best Local Similarity 100.0%; Pred. No. 0.04;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ELGPGGRPPPC 12

Db 1 ELGPGGRPPPC 12

RESULT 4

US-09-700-378-5

Sequence 5, Application US/09700378

GENERAL INFORMATION:

APPLICANT: Gevas, Philip

APPLICANT: Stephen, Grimes

APPLICANT: Karr, Stephen

APPLICANT: Michaeli, Dov

TITLE OF INVENTION: Method for the Treatment of Gastroesophageal Reflux Disease

FILE REFERENCE: ACG1USA

CURRENT APPLICATION NUMBER: US/09/700,378

CURRENT FILING DATE: 2000-11-14

PRIOR APPLICATION NUMBER: PCT/US99/10734

PRIOR FILING DATE: 1999-05-14

PRIOR APPLICATION NUMBER: 60/085,610

PRIOR FILING DATE: 1998-05-15

NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn version 3.0

SEQ ID NO 5

LENGTH: 12

TYPE: PRT

ORGANISM: human

FEATURE:

NAME/KEY: MOD\_RES

LOCATION: (1)-(1)

OTHER INFORMATION: pyroglutamic acid

US-09-700-378-5

Query Match 100.0%; Score 75; DB 21; Length 12;

Best Local Similarity 100.0%; Pred. No. 0.04;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ELGPGGRPPPC 12

Db 1 ELGPGGRPPPC 12

RESULT 5

PCT-US94-04832A-73

Sequence 73, Application PC/TUS9404832A

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

STREET: 400 Garden City Plaza

CITY: Garden City

STATE: NY

COUNTRY: US

ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/04832A

FILING DATE: 13-APR-1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME:

```

;
;   REGISTRATION NUMBER:
;   REFERENCE/DOCKET NUMBER:
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (516)742-4343
;   TELEFAX: (516)742-4366
;   INFORMATION FOR SEQ ID NO: 73:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 11 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   PCT-US94-04832A-73

Query Match
Best Local Similarity 74.7%; Score 56; DB 1; Length 11;
Pred. No. 6.7;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ELGPGGRPPP 11
Db 1 QLGPGGPPPP 11

RESULT 6
US-08-488-320A-73
; Sequence 73, Application US/08488320A
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic Peptides which Contain LHRH
; TITLE OF INVENTION: And A Helper T-Cell Epitope For Treatment Of Prostate Cancer
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,320A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/488,351
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: 08/446,692
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: 08/229,275
; FILING DATE: 14-APR-1994
; APPLICATION NUMBER: 08/057,166
; FILING DATE: 27-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C. H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4146USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)758-4800
; TELEFAX: (212)751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-488-320A-73
```

```

Query Match
Best Local Similarity 74.7%; Score 56; DB 8; Length 11;
Pred. No. 6.7;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ELGPGGRPPP 11
Db 1 QLGPGGPPPP 11

RESULT 7
PCT-US01-14827-14717
; Sequence 14717, Application PC/TUS0114827
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
; CURRENT APPLICATION NUMBER: PCT/US01/14827
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 16102
; SOFTWARE: Custom
; SEQ ID NO 14717
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(160)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
PCT-US01-14827-14717

Query Match
Best Local Similarity 74.7%; Score 56; DB 1; Length 160;
Pred. No. 76;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PGRPPPPC 12
Db 138 PGRPPPPC 146

RESULT 8
PCT-US01-08656-6637
; Sequence 6637, Application PC/TUS0108656
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-066
; CURRENT APPLICATION NUMBER: PCT/US01/08656
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 09/522,929
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: 09/770,160
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 10994
; SOFTWARE: Custom
; SEQ ID NO 6637
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(139)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
PCT-US01-08656-6637

Query Match
Best Local Similarity 73.3%; Score 55; DB 1; Length 139;
Pred. No. 88;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ELGPGGRPPP 11
Db 1 ELGPGGRPPP 11
```

DB 29 QISPOGRPPPP 39

RESULT 9  
PCT-US01-08656-6639  
Sequence 6639, Application PC/TUS0108656  
GENERAL INFORMATION:  
APPLICANT: Hysq, Inc  
TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES  
FILE REFERENCE: 21272-066  
CURRENT APPLICATION NUMBER: PCT/US01/08656  
CURRENT FILING DATE: 2001-04-16  
PRIOR APPLICATION NUMBER: 09/522,929  
PRIOR FILING DATE: 2000-04-18  
PRIOR APPLICATION NUMBER: 09/770,160  
PRIOR FILING DATE: 2001-01-26  
NUMBER OF SEQ ID NOS: 10994  
SOFTWARE: Custom  
SEQ ID NO 6639  
LENGTH: 169  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: (1)...(169)  
OTHER INFORMATION: Xaa = X or \* as defined in Table 2  
PCT-US01-08656-6639

Query Match  
Best Local Similarity 73.3%; Score 55; DB 1; Length 169;  
Best Local Similarity 72.7%; Pred. No. 1.1e+02;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ELPOGRPPPP 11  
DB 14 QISPOGRPPPP 24

RESULT 10  
US-09-758-444-470  
Sequence 470, Application US/09758444  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PM032  
CURRENT APPLICATION NUMBER: US/09/758,444  
CURRENT FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
NUMBER OF SEQ ID NOS: 638  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 470  
LENGTH: 185  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (72)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (73)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (74)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (75)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (83)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE

LOCATION: (84)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (105)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-758-444-470

Query Match  
Best Local Similarity 69.3%; Score 52; DB 21; Length 185;  
Best Local Similarity 80.0%; Pred. No. 2.6e+02;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GPVGRPPPPC 12  
DB 37 GPVGRPPPPC 46

RESULT 11  
US-09-758-444-476  
Sequence 476, Application US/09758444  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PM032  
CURRENT APPLICATION NUMBER: US/09/758,444  
CURRENT FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
NUMBER OF SEQ ID NOS: 638  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 476  
LENGTH: 323  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (77)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (86)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (134)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (198)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (290)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (293)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (319)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-758-444-476

Query Match  
Best Local Similarity 69.3%; Score 52; DB 21; Length 323;  
Best Local Similarity 80.0%; Pred. No. 4.3e+02;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GPVGRPPPPC 12  
DB 32 GPVGRPPPPC 41

RESULT 12  
US-10-212-160-476

```
; Sequence 476, Application US/10211160
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM032C1N
; CURRENT APPLICATION NUMBER: US/10/212,160
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: 09/758,444
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 476
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (86)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (134)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (198)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (290)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (293)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-212-160-476

Query Match
Best Local Similarity 69.3%; Score 52; DB 26; Length 333;
Pred. No. 4.3e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 3 GPGGRPPPC 12
    11111111
Db 32 GPGGRPPPC 41
```

```
RESULT 13
; Sequence 10407, Application US/10015127
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Sphingomonas elodea genome sequences and uses thereof
; FILE REFERENCE: 38-10(15806)B
; CURRENT APPLICATION NUMBER: US/10/015,127
; CURRENT FILING DATE: 2001-10-29
```

```
; PRIOR APPLICATION NUMBER: US 60/252,455
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 14357
; SEQ ID NO 10407
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Sphingomonas elodea
; US-10-015-127-10407
```

```
Query Match
Best Local Similarity 68.7%; Score 51.5; DB 24; Length 172;
Pred. No. 2.8e+02;
Matches 9; Conservative 1; Mismatches 1; Indels 1; Gaps 1;
```

```
OY 2 LGPGRRP-PPC 12
    11111111
Db 126 VGPSGRPPVPPC 137
```

```
RESULT 14
; Sequence 962, Application PC/TUS9922855
; GENERAL INFORMATION:
; APPLICANT: Ceres, Inc.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding
; FILE REFERENCE: 2750-0567(PC)
; CURRENT APPLICATION NUMBER: PCT/US99/22855
; CURRENT FILING DATE: 1999-10-05
; NUMBER OF SEQ ID NOS: 3978
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 962
; LENGTH: 144
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: any n or Xaa = unknown
; OTHER INFORMATION: LOCATION 1..144 / Ceres Seq. ID 1686426
; PCT-US99-22855-962
```

```
Query Match
Best Local Similarity 68.0%; Score 51; DB 1; Length 144;
Pred. No. 2.7e+02;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 1 ELGPRRPPPP 11
    11111111
Db 54 QGPRRPPPP 64
```

```
RESULT 15
; Sequence 4290, Application US/09451320
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai
; APPLICANT: CHEN, Xianfeng
; APPLICANT: SUBRAMANIAN, Gopalakrishnan
; APPLICANT: ZHENG, Liangsheng
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS 42-59, 62-66, 68, 81
; FILE REFERENCE: 2750-0662P
; CURRENT APPLICATION NUMBER: US/09/451,320
; CURRENT FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 6998
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4290
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: any n or Xaa = unknown
; US-09-451-320-4290

Query Match
Best Local Similarity 68.0%; Score 51; DB 18; Length 222;
Pred. No. 4e+02;
```

	Matches	8;	Conservative	2;	Mismatches	1;	Indels	0;	Gaps	0;
QY	1	ELGPGGRPPPP	11							
		:   :								
Db	132	QLGPRRPPPP	142							

Search completed: February 5, 2003, 09:55:40  
Job time : 157 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2003, 09:48:01 ; Search time 19.4286 Seconds  
(without alignments)  
50.530 Million cell updates/sec

Title: US-09-700-329-2

Perfect score: 75

Sequence: 1 ELGPGRRPPPC 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412725 segs, 81809943 residues

Total number of hits satisfying chosen parameters: 412725

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New\*

1: /cgn2\_6/ptodata/1/paa/PC7\_NEM\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEM\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEM\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEM\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEM\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEM\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US60\_NEM\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75	100.0	12	US-10-314-057-5	Sequence 5, Appl1
2	50	66.7	163	US-10-154-678-58	Sequence 58, Appl1
3	50	66.7	265	US-09-724-676-88534	Sequence 88534, A
4	50	66.7	265	US-09-724-676A-88534	Sequence 88534, A
5	50	66.7	319	US-09-724-676-88519	Sequence 88519, A
6	50	66.7	319	US-09-724-676-88523	Sequence 88523, A
7	50	66.7	319	US-09-724-676-88527	Sequence 88527, A
8	50	66.7	319	US-09-724-676-88532	Sequence 88532, A
9	50	66.7	319	US-09-724-676-88535	Sequence 88535, A
10	50	66.7	319	US-09-724-676-88539	Sequence 88539, A
11	50	66.7	319	US-09-724-676A-88519	Sequence 88519, A
12	50	66.7	319	US-09-724-676A-88523	Sequence 88523, A
13	50	66.7	319	US-09-724-676A-88527	Sequence 88527, A
14	50	66.7	319	US-09-724-676A-88532	Sequence 88532, A
15	50	66.7	319	US-09-724-676A-88535	Sequence 88535, A
16	50	66.7	319	US-09-724-676A-88539	Sequence 88539, A
17	50	66.7	332	US-09-724-676-88517	Sequence 88517, A
18	50	66.7	332	US-09-724-676-88521	Sequence 88521, A
19	50	66.7	332	US-09-724-676-88525	Sequence 88525, A
20	50	66.7	332	US-09-724-676-88530	Sequence 88530, A
21	50	66.7	332	US-09-724-676-88537	Sequence 88537, A
22	50	66.7	332	US-09-724-676A-88541	Sequence 88541, A
23	50	66.7	332	US-09-724-676A-88517	Sequence 88517, A
24	50	66.7	332	US-09-724-676A-88521	Sequence 88521, A
25	50	66.7	332	US-09-724-676A-88525	Sequence 88525, A
26	50	66.7	332	US-09-724-676A-88530	Sequence 88530, A

27	50	66.7	332	US-09-724-676A-88537	Sequence 88537, A
28	50	66.7	332	US-09-724-676A-88541	Sequence 88541, A
29	50	66.7	342	US-09-724-676-88518	Sequence 88518, A
30	50	66.7	342	US-09-724-676-88522	Sequence 88522, A
31	50	66.7	342	US-09-724-676-88526	Sequence 88526, A
32	50	66.7	342	US-09-724-676-88528	Sequence 88528, A
33	50	66.7	342	US-09-724-676-88531	Sequence 88531, A
34	50	66.7	342	US-09-724-676-88538	Sequence 88538, A
35	50	66.7	342	US-09-724-676A-88518	Sequence 88518, A
36	50	66.7	342	US-09-724-676A-88522	Sequence 88522, A
37	50	66.7	342	US-09-724-676A-88526	Sequence 88526, A
38	50	66.7	342	US-09-724-676A-88528	Sequence 88528, A
39	50	66.7	342	US-09-724-676A-88531	Sequence 88531, A
40	50	66.7	342	US-09-724-676A-88538	Sequence 88538, A
41	50	66.7	343	US-09-724-676-88533	Sequence 88533, A
42	50	66.7	343	US-09-724-676A-88533	Sequence 88533, A
43	50	66.7	355	US-09-724-676-88520	Sequence 88520, A
44	50	66.7	355	US-09-724-676-88524	Sequence 88524, A
45	50	66.7	355	US-09-724-676-88529	Sequence 88529, A

## ALIGNMENTS

```

RESULT 1
US-10-314-057-5
: Sequence 5, Application US/10314057
: GENERAL INFORMATION:
: APPLICANT: Geyas, Philip
: APPLICANT: Stephen, Grimes
: APPLICANT: Kier, Stephen
: APPLICANT: Michael, Dov
: TITLE OR INVENTION: Method for the Treatment of Gastroesophageal Reflux Disease
: FILE REFERENCE: REG15A
: CURRENT APPLICATION NUMBER: US/10/314,057
: CURRENT FILING DATE: 2002-12-06
: PRIOR APPLICATION NUMBER: US/09/700,378
: PRIOR FILING DATE: 2000-11-14
: PRIOR APPLICATION NUMBER: PCT/US99/10734
: PRIOR FILING DATE: 1999-05-14
: PRIOR APPLICATION NUMBER: 60/085,610
: PRIOR FILING DATE: 1998-05-15
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 5
: LENGTH: 12
: TYPE: PRT
: ORGANISM: human
: FEATURE:
: NAME/KEY: MOD_RES
: LOCATION: (1)..(1)
: OTHER INFORMATION: pyroglutamic acid
US-10-314-057-5
Query Match 100.0%; Score 75; DB 6; Length 12;
Best Local Similarity 100.0%; Pred No. 0.0038;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ELGPGRRPPPC 12
Db 1 ELGPGRRPPPC 12
RESULT 2
US-10-154-678-58
: Sequence 58, Application US/10154678
: GENERAL INFORMATION:
: APPLICANT: Benjamin, Stephanie
: APPLICANT: Tanaka, Hiroaki
: TITLE OR INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
: FILE REFERENCE: 182.051.REG
: CURRENT APPLICATION NUMBER: US/10/154,678
: CURRENT FILING DATE: 2002-10-15

```

; PRIOR APPLICATION NUMBER: US 09/924,340  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: Jpatent  
; SEQ ID NO 58  
; LENGTH: 163  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-154-678-58

Query Match 66.7%; Score 50; DB 6; Length 163;  
Best Local Similarity 80.0%; Pred. No. 30;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 LGPGRRPPP 11  
Db 43 MGPGRRPPP 52

RESULT 3  
US-09-724-676-88534  
; Sequence 88534, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; NUMBER OF SEQ ID NOS: 2000-11-28  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 88534  
; LENGTH: 265  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-88534

Query Match 66.7%; Score 50; DB 5; Length 265;  
Best Local Similarity 88.9%; Pred. No. 43;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GPGRRPPP 11  
Db 162 GPGRRPPP 170

RESULT 4  
US-09-724-676A-88534  
; Sequence 88534, Application US/09724676A  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676A  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 88534  
; LENGTH: 265  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676A-88534

Query Match 66.7%; Score 50; DB 5; Length 265;  
Best Local Similarity 88.9%; Pred. No. 43;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GPGRRPPP 11  
Db 162 GPGRRPPP 170

RESULT 5  
US-09-724-676-88519  
; Sequence 88519, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 88519  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-88519

Query Match 66.7%; Score 50; DB 5; Length 319;  
Best Local Similarity 88.9%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GPGRRPPP 11  
Db 162 GPGRRPPP 170

RESULT 6  
US-09-724-676-88523  
; Sequence 88523, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 88523  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-88523

Query Match 66.7%; Score 50; DB 5; Length 319;  
Best Local Similarity 88.9%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GPGRRPPP 11  
Db 162 GPGRRPPP 170

RESULT 7  
US-09-724-676-88527  
; Sequence 88527, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 88527  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-09-724-676-88527

Query Match 66.7%; Score 50; DB 5; Length 319;  
Best Local Similarity 88.9%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GPGGRPPP 11  
|||

Db 162 GPGGRPPP 170

RESULT 8

US-09-724-676-88532  
; Sequence 88532, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 88532  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-88532

Query Match 66.7%; Score 50; DB 5; Length 319;  
Best Local Similarity 88.9%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GPGGRPPP 11  
|||

Db 162 GPGGRPPP 170

RESULT 9

US-09-724-676-88535  
; Sequence 88535, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 88535  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-88535

Query Match 66.7%; Score 50; DB 5; Length 319;  
Best Local Similarity 88.9%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GPGGRPPP 11  
|||

Db 162 GPGGRPPP 170

RESULT 10

US-09-724-676-88539  
; Sequence 88539, Application US/09724676  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 88539  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-88539

Query Match 66.7%; Score 50; DB 5; Length 319;  
Best Local Similarity 88.9%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GPGGRPPP 11  
|||

Db 162 GPGGRPPP 170

RESULT 11

US-09-724-676A-88519  
; Sequence 88519, Application US/09724676A  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676A  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 88519  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676A-88519

Query Match 66.7%; Score 50; DB 5; Length 319;  
Best Local Similarity 88.9%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GPGGRPPP 11  
|||

Db 162 GPGGRPPP 170

RESULT 12

US-09-724-676A-88523  
; Sequence 88523, Application US/09724676A  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676A  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 88523  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676A-88523

Query Match 66.7%; Score 50; DB 5; Length 319;  
Best Local Similarity 88.9%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GPGGRPPP 11  
|||

Db 162 GPGGRPPP 170

RESULT 13

US-09-724-676A-88527  
; Sequence 88527, Application US/09724676A  
; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD

Search completed: February 5, 2003, 09:50:58  
 Job time : 19.4286 secs

```

: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724, 676A
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 88527
: LENGTH: 319
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676A-88527

```

```

Query Match          66.7%; Score 50; DB 5; Length 319;
Best Local Similarity 88.9%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 3 GPGRRPPP 11
DB 162 GPGRRPPP 170

```

```

RESULT 14
US-09-724-676A-88532
: Sequence 88532, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724, 676A
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 88532
: LENGTH: 319
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676A-88532

```

```

Query Match          66.7%; Score 50; DB 5; Length 319;
Best Local Similarity 88.9%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 3 GPGRRPPP 11
DB 162 GPGRRPPP 170

```

```

RESULT 15
US-09-724-676A-88535
: Sequence 88535, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724, 676A
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 88535
: LENGTH: 319
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676A-88535

```

```

Query Match          66.7%; Score 50; DB 5; Length 319;
Best Local Similarity 88.9%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 3 GPGRRPPP 11
DB 162 GPGRRPPP 170

```